Controls and counterfactual information in agro-ecological investment

David Duncan and Paul Reich

Key lessons

• Appropriate contrasts, such as controls and counterfactual data, are fundamental to sound interpretation of the effectiveness of agri-environment schemes.
• Such contrasts are rarely included in evaluations of Australian agri-environment schemes for a range of reasons, including logistical constraints.
• Different kinds of contrasts exist that permit different kinds of inference about program effectiveness.
• Effective evaluation incorporating sampling counterfactual data need not cost more than is currently expended on monitoring and evaluation.
• Every scheme should explicitly include counterfactual thinking in evaluation plans, even if there is no intention to monitor.
Figure 19.1: A riparian zone near Euroa, Victoria, that has been fenced and replanted with a mix of native species. Evaluating the benefits of projects such as these requires an understanding of what would have happened in the absence of the project — the counterfactual.

Source: Photo by Paul Reich.
Introduction

Despite the large amounts of money invested in agri-environment schemes in Australia (Hajkowicz 2009), there remains high uncertainty about the magnitude of expected environmental benefit. Sophisticated approaches for dealing with uncertainty are now routinely adopted in systematic conservation planning processes (e.g. prioritisation, optimisation, Sarkar et al. 2006). Unfortunately, these advances have not been matched in the evaluation, learning, and improvement part of the decision cycle, where conservation and environmental management lag behind other complex domains such as social policy and medicine (Stem et al. 2005; Ferraro 2009; Field et al. 2007).

Reporting of management performance (activity and outputs, sensu Mascia et al. 2014) in agri-environment schemes has itself been patchy, but direct demonstrations of the impact of intervention, that is, the difference in change between intervention and non-intervention sites, are exceedingly rare (Margoluis et al. 2009; Ferraro and Pattanayak 2006; but see Hale et al. 2011; Lindenmayer et al. 2012). As we discuss in this chapter, this so-called counterfactual evidence (derived from control sites without intervention) is fundamentally important for meaningful evaluation of agri-environmental investment schemes. We outline the difficulties that investment in environmental change poses for management experiments, and suggest positive ways of addressing those difficulties.

Our focus is the type of evaluation questions identified by Mascia et al. (2014) as ‘impact evaluation’ where the interpretation logically demands a counterfactual contrast (Ferraro 2009). For example, what is the impact of livestock exclusion from remnant vegetation on the abundance of sensitive native herbs?

In most cases, the effectiveness of work funded by agri-environment schemes is evaluated using post-hoc, space-for-time substitution surveys (e.g. Prober et al. 2011; Read et al. 2011). To accept the implied effect, we assume that the sites were equivalent at some point in the past and that the difference is due to the funded intervention, but we cannot be certain of this. Particularly in agricultural settings, many confounding factors exist that could inflate or obscure the effects of the intervention. Also, we ignore the possibility of influence of interactions between, for example, climatic regime and the
intervention. More powerful conclusions about the effectiveness of changed management can be made when data are collected through time, for the period of the intervention.

Responses to large-scale interventions have been evaluated where sufficient long-time series data were available to allow change point analysis (i.e. the detection of a change in a time series, Box and Tiao 1975; Stewart-Oaten and Bence 2001; Thomson et al. 2010; Stoffels and Weatherman 2014). These examples have no direct counterfactual sampling option because the subject or events cannot be replicated meaningfully (e.g. matching an entire river, estuary, city, etc.). Instead the interpretation relies on consideration of whether the observed change could have been generated independently of the intervention. This is one example of model-based counterfactual inference, where an expected alternative under no intervention may be credibly argued with reference to weight of data accumulated prior to the intervention.

By contrast, the investments in agri-environmental schemes are usually conceptually and practically replicable, although we acknowledge that important constraints exist. Time series analysis based on intervention sites are usually impossible because the locations where interventions will occur are rarely known for long enough in advance to enable adequate data collection, and the interventions themselves are relatively short term. This means that direct counterfactual inference must come from simultaneous sampling of control and intervention sites.

Why are controls needed to estimate the impact of agri-environmental schemes?

By making payments, agri-investment schemes seek to change the status quo. The basic assumption is that those sites or landscapes where investment (intervention) is made will have a different future to sites where no intervention occurs. When change is estimated only from intervention sites, there is an implicit assumption that non-funded sites will not change (e.g. Figure 19.2a). While this is one plausible scenario, there are others that should be considered. Counterfactual information from control sites allows us to weigh rival interpretations of the outcomes of interventions (Ferraro 2009).
Figure 19.2: Simplified representation of mean agri-environment program outcomes, contrasting mean change over the course of investment in intervention sites (solid line) against mean background change (dotted lines, panels a–f), and mean change in reference sites (dashed lines, panels d–f).

Source: Authors’ research.

For example, the most recent State of the Environment Committee Report (2011) concluded that native ecosystems on private land are mostly in decline. This suggests that agri-environment schemes could be considered successful even if managed sites show a reduced decline in condition compared with non-intervention sites (e.g. Figure 19.2b).

Another possibility is that positive changes in the extent of native ecosystems, consistent with the objectives of agri-environment schemes, are occurring spontaneously due to declining extent of agricultural production and land use changes (Kyle and Duncan 2012; Geddes et al. 2011). Government investment in agri-environment schemes may have only marginal benefit over the improving background trend (e.g. Figure 19.2c).

Counterfactual data from control sites — or at minimum a coherent conceptual model of the presumed fate of control sites (see the final option of Table 19.1) — are required to interpret responses measured at intervention sites and appropriately evaluate the success or otherwise of a given agri-environmental investment scheme. As Figure 19.2d–f
illustrates, an additional contrast against reference sites (the desired state) can provide valuable additional insight into the relative performance of interventions against non-intervention (Downes et al. 2002; Coffman et al. 2014).

There are different types of controls and contrasts

In the context of agri-environment schemes, distinct options for control sites exist that permit subtle differences in the inference possible (Table 19.1). It is important to think about what these options mean, and the limits to their interpretation. Here we present a range of different control options available for comparison, and assess each against its likely relative value with regards (1) strength of inference for a given sample size, (2) transferability of the inference beyond the specific program, (3) contribution to a causal understanding, and (4) capacity to accommodate multiple response variables.

The most efficient learning scenario is to construct a management experiment where there is potential to select sites based on a management question, and then randomly allocate sites to treatment and non-treatment classes. This scenario, for a given sample size, offers the strongest chance to learn about what works, where, and why. However, it can be hard (and often impossible) to get support and sufficient sample sizes for experimentation on large scales, particularly involving public money on private land. In the past, many Australian states and the CSIRO had access to publicly owned production land where demonstration farming and experimentation took place for agricultural productivity. Such holdings may offer a cheaper and more secure opportunity for the Australian Government to learn about the effectiveness of interventions in comparison to building management experiments into the implementation of agri-environment schemes.
243

19. CONTROLS AND COUNTERFACTUAL INFORMATION IN AGRO-ECOLOGICAL INVESTMENT

Table 19.1: Attributes and constraints of distinct counterfactual and contrast scenarios, spanning tailored management experiments where the strongest inference might be anticipated through to model-based calculation of impact, for which the major constraint is the paucity of available evidence from the types of sampling listed higher in the table.

<table>
<thead>
<tr>
<th>Type of control / contrast</th>
<th>Attributes relating to inference</th>
<th>Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strength</td>
<td>Transferability</td>
</tr>
<tr>
<td>Management experiment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paired site on winning bid site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unsuccessful bid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random sample</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space for time substitution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference (contrast against desired state)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model based counterfactual</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors’ research.

A more likely route to ensure adequate sample sizes is to build quantitative evaluation into major investment programs. In this instance, the researcher has to reactively match control sites to sites selected by the funding agency. Lindenmayer et al. (2012), for example, have established a major evaluation of the impact of 10–15-year grazing management agreements in the Box Gum Grassy Woodlands on native flora and fauna, comparing treatment sites with control sites located on the property of winning bidders in the auction program. This approach takes advantage of the convenience of an established relationship with the participating landholder, and should reduce sources of random variation, such as spatial variation in environmental factors, and farm level management factors (both historical and contemporary). These benefits permit a relatively robust comparison of the difference in change between treatments and sites under pre-existing management, although some limitations exist (see below).

Duncan and Vesk (2013) suggested that unsuccessful bidders for agri-investment payments could be a potential source of control sites for successful bids. These sites have the advantage of being assessed
as part of the bidding process, so landholders have been engaged and some data may be available to guide their inclusion. However, greater random difference due to spatial effects and past management might be expected, so the comparison would be expected to be noisier, and the inference weaker, for a given sample size compared to a matched control on a winning bid. In both of the preceding cases, it would be hard to match the starting condition of funded sites with control sites, as the selection process is designed to favour the best-quality sites available for intervention.

The potential downside of locating control sites on the property of a paid participant is that the existence of the treatment, and the accompanying negotiation, may bias a participant’s approach to management. After all, permanent behavioural and attitudinal change is one objective of government investment in agri-investment schemes, rather than merely switching on favourable management for the duration of a contract. Lindenmayer et al. (2012) expressed confidence the management of control sites was unaffected by the management of the paired treatment, but elsewhere involvement in auctions has been shown to influence the way landholders manage non-intervention areas (Windle et al. 2009).

These comparisons do not enable change in treatment sites to be compared against the way the average site is managed, but rather against the way a landholder positively disposed to conservation programs might manage their land. The estimates of background change we obtain from controls in agri-environment schemes are therefore unlikely to represent the average trend from the broader landscape, which may constitute a more desirable impact statement for program managers. We might ideally like to randomly sample appropriate controls for funded treatment units from the broader landscape, however, in practice problems of selection bias remain in those that choose to allow their properties to be visited and sites sampled. An estimate of variation associated with a randomly selected group of control sites would also require a larger sample size.

The minimal option should be an explicit, model-based counterfactual comparison (see final option of Table 19.1). A simple version might involve sampling intervention sites, with the amount of change being subsequently claimed as impact against a clear justification for what control sites are expected to do. The most important change from the
way evaluation is typically conducted at present is that the model of control site behaviour must be made explicit and well justified. We are not aware of any examples of this kind of practice in action. More commonly, change data from intervention sites is claimed as an impact, with no declaration about what is presumed to be happening under business-as-usual scenarios.

One can imagine further sophistications of this model-based approach, where relevant quantitative data on background change under business-as-usual scenarios can be used to simulate the likely behaviour of control scenarios in evaluation. However, this possibility will not generally be available or particularly compelling until relevant data is accumulated from the more conventional approaches higher up in Table 19.1.

**Why are controls and counterfactual contrasts so rarely obtained?**

There are challenges to be overcome in identifying, negotiating, funding, and interpreting appropriate counterfactual data to use in evaluating agri-environmental schemes (e.g. Ferraro 2009; Kibler 2011). Perhaps due to these challenges, program specific or generic monitoring and evaluation advice relevant to agri-environment schemes may be released without any mention of their importance, or considerations for sampling controls or counterfactual data (e.g. the Conservation Measures Partnership 2013; DSEWPaC 2013). Such omissions contribute to evaluation design that does not explicitly discuss how counterfactual evidence will be gathered, inferred, or done without. We agree with Ferraro (2009) that evaluation plans should at least demonstrate counterfactual thinking.

**Cost of sampling control sites**

One factor that surely limits agency support for sampling control sites is cost, usually conceived as an increase in the total monitoring budget. A simplistic assumption might be that including (paired) controls in the sampling design would double the cost allocated to monitoring, further diminishing the amount assigned to action. However, one of the primary reasons to monitor is to demonstrate impact and learn
about treatment effectiveness (e.g. this program has increased the occupancy of woodland birds by X per cent over a background trend of Y per cent). Therefore, it can readily be demonstrated that control sampling is fundamental to all cost-effective designs, as no amount of monitoring without controls can support the required inference.

Once the desired result statements are clearly defined, one can simulate the data collection and analysis in advance to examine which sampling scenarios are mostly likely to cost-effectively deliver that result. In reality, strategic sampling for impact evaluation, by including a subset of treatments and controls, could probably be achieved for a similar cost to that typically spent on monitoring programs which have failed to generate strong insight about the effectiveness of interventions.

**Funding models and process**

The way in which agencies tend to allocate and deliver funding for agri-environmental schemes, at both program and project level, can make it difficult to design and implement strong quantitative evaluation. For example, projects may be awarded funds concurrent with, or even before, the design of evaluation, making it impossible to obtain pre-intervention data from intervention sites, let alone control sites.

**Lack of clarity of objectives and process model hinders evaluation design**

The failure to make explicit program objectives and a conceptual or process model of cause and effect hinders monitoring of any sort (Field et al. 2007), and compounds the difficulty in identifying an appropriate control in program evaluation. The objectives and assumptions in the conceptual model of cause and effect should indicate what sort of trajectories and effect sizes to expect, and also guide the selection of a suitable control.

Some investment objectives are particularly difficult to control for. While controls should be achievable for site-scale interventions and responses (e.g. Lindenmayer et al. 2012; Hale et al. 2014), the greater the spatial scale of the program objective (e.g. landscape connectivity), or the greater the number of links in the causal chain between source
and impact (e.g. changing agricultural land use to reduce oceanic hypoxia, Rabotyagov et al. 2014), the harder it may be to identify appropriate controls.

**Horses for courses in counterfactual data collection**

We intuitively think of the ideal control site as an independent site, closely matched to our treatment. Constructive solutions could be identified by focusing instead on the specific counterfactual requirements for response variables in the conceptual model of cause and effect. Within a given investment program, this may imply different data gathered at different scales or indeed locations. An appropriate control for measuring aquatic responses should differ from terrestrial plant responses, which could be different again for faunal responses.

Consider investment in restoration of riparian corridors. The impact of the investment on terrestrial vegetation might be well accommodated by a fenceline contrast (i.e. comparison of management regime either side of a fence), whereas the control for aquatic responses may be best placed some distance upstream of the treatment to maximise independence owing to the directional movement of water and its constituents. For occupancy responses of mobile biota, direction may not matter, but distance between sites may be important to achieve requisite independence.

**What is required to improve our understanding?**

The design of evaluation for all agri-environmental schemes should explicitly include counterfactual thinking (Ferraro 2009). In theory, this thinking should be represented in program logic diagrams or conceptual models that set out the expected difference in outcome comparing intervention and no-intervention (e.g. model-based counterfactual, Table 19.1). Larger schemes should produce well designed and resourced quantitative evaluation, linked to those models.

*Be realistic.* High rigour generally means less replication and reduced coverage of important contexts or covariates. A strategic mix of observational and experimental studies that explicitly complement and reference each other are required.
Large and small agri-environment schemes should do the best with what is available, including supporting post hoc comparisons, and using simulation models and scenario analyses. All techniques that can help make the most of existing data will remain important, and sound evaluation design should inform and guide data requirements. However, none of these fallback options excuse the persistent failure to conduct robust evaluation of agri-environment programs, including obtaining counterfactual data.

Be upfront about limitations to interpretation. Where less than ideal evaluation and assessment takes place, it is important to clearly state the limits to interpretation. For example, Duncan and Vesk (2013) estimated a substantial reduction in weed cover in sites funded by Victoria’s BushTender program comparing before and after data from intervention sites. However, due to the lack of control sites, they explicitly cautioned that the observed changes were just as plausibly attributable to sustained drought.

Synthesise and disseminate. There are major programs beginning to establish rolling synopses of evidence of effects of different agri-environmental interventions (Dicks et al. 2013; Pullin and Knight 2009), including studies containing counterfactual evidence. Those synopses are tailored to the implementation context of northern and western Europe, so Australia should expect to support its own version, given our environmental, cultural and land use history and pattern.

Conclusions and recommendations

The current forms of monitoring and reporting (e.g. MERI — Monitoring, Evaluation Reporting, Improvement — Australian Government Land and Coasts 2009) undertaken in Australia have a valid role in the delivery and evaluation of agri-environmental schemes, but there is an urgent need to translate rhetoric into disciplined practice in quantifying environmental impact. However, our current systems routinely deliver poorly designed data collection activities, the results of which are scarcely, if ever, analysed and publicised.

Considerable coordination and nuance may be required to obtain inference about the impact of interventions in a cost-effective manner. For example, counterfactual data for interventions may be sourced at different spatial and temporal scales, as defined by the conceptual
model relationship between treatment and response. It is likely that no evaluation program will encompass all elements and scales of space and time, but every program should be expected to make a coherent statement about effectiveness that includes an explicit contrast with a non-intervention scenario.

Done well, effective evaluation incorporating counterfactual data need not cost more than is currently expended on monitoring and evaluation. Importantly, even though not all programs will undertake such sampling, all should explicitly represent counterfactual thinking in MERI plans and program design. In addition to an immense literature relevant to setting objectives for agri-environment schemes, we offer the following checklist for evaluating whether a MERI plan for an agri-environment scheme has met minimum requirements:

1. The management behaviour or resource trend that funded treatments are intended to address, ameliorate, or reverse should be specified, in its appropriate spatial and temporal context.
2. The counterfactual prognosis (in terms of averages and some indication of variation) should be specified for the term of the funded treatments, and beyond, according to the definition of 1.
3. Elements 1 and 2 should be expressed in a manner that conveys the degree of certainty and scientific consensus, regarding averages and sources of variation, so that MERI programs that will guide field data collection are designed for maximum benefit.

Acknowledgements

We thank Steve Sinclair for discussions and suggestions that improved an earlier draft. Dean Ansell and anonymous reviewers provided ideas and suggestions that further improved our final manuscript.

References


