

Data Analysis Plan

The Effect of Wildflower Strips on Pollinator Species Richness

Ed Harris

2026-01-13

Table of contents

1 Title: The Effect of Wildflower Strips on Pollinator Species Richness in UK Agricultural Landscapes	1
2 Background & Objectives	1
3 Data and Statistical Methods	3
4 Outputs, Interpretation & Reporting	12
5 Reproducibility & Amendments	14
6 Resources	16
7 Appendix A: Power Analysis Calculations	17

1 Title: The Effect of Wildflower Strips on Pollinator Species Richness in UK Agricultural Landscapes

2 Background & Objectives

2.1 Background

Pollinator decline represents a critical threat to global biodiversity and ecosystem services, with significant implications for agricultural productivity and food security. In response to these declines, conservation strategies such as wildflower strips have been implemented across agricultural landscapes to enhance pollinator habitat and support biodiversity. Wildflower strips are semi-natural habitats established within or adjacent to agricultural fields, designed to provide floral resources and nesting sites for pollinators throughout the growing season.

Meta-analyses have demonstrated that flower strips and hedgerows can effectively enhance pest control and pollination services, though effects on crop yield vary (Albrecht et al., 2020). However, the magnitude of these effects, particularly for pollinator species richness, remains variable across different contexts and regions. Research across four European countries has shown that both local floral resources (within wildflower strips) and landscape-level floral resources influence the effectiveness of wildflower strips for supporting wild bee communities (Scheper et al., 2015). These findings highlight the importance of understanding context-dependent effects and the need for robust quantification of effect sizes.

This analysis will examine the effect of wildflower strips on pollinator species richness using a pairwise experimental design across multiple UK farms. The study employs a nested sampling structure with multiple samples per farm, allowing for the estimation of both within-farm and between-farm variation. Given the nested design and expected small effect size, careful attention to statistical power and appropriate modeling approaches is essential.

Important constraints: This is an observational study using existing wildflower strip projects on working farms. Farms are not randomly assigned to treatment conditions, but rather selected based on participation in wildflower strip programs. Within each farm, pairwise comparisons are made between fields with wildflower strips (treatment) and adjacent control fields without strips. Sampling occurs weekly during the peak pollinator season (approximately 12-16 weeks) over two consecutive years, providing temporal replication.

2.2 Objectives

Primary objective(s):

1. To estimate the effect of wildflower strips on pollinator species richness, quantified as the difference in mean species richness between treatment fields (with wildflower strips) and control fields (without strips) within the same farms.
2. To quantify the magnitude of this effect using standardized effect size measures (Cohen's d), with an expected small effect size of approximately 0.2.

Secondary / exploratory objective(s):

1. To examine temporal variation in the effect of wildflower strips across sampling weeks and between years.

2. To explore whether farm-level characteristics (e.g., farm size, crop type, landscape context) moderate the effectiveness of wildflower strips.
3. To assess the consistency of effects across different pollinator functional groups (e.g., bees, hoverflies, butterflies) if sufficient taxonomic resolution is available.

3 Data and Statistical Methods

3.1 Data sources and structure

Data source(s) and how data were generated:

Data will be collected through standardized field surveys on UK farms participating in wildflower strip conservation programs. Pollinator sampling will be conducted using pan traps, a widely used and standardized method for assessing pollinator diversity. Pan traps are passive sampling devices that effectively capture a broad range of pollinator taxa, providing consistent and comparable data across sites and time periods.

Independent unit(s) of analysis:

The primary unit of analysis is the individual sampling event (pan trap collection), nested within fields (treatment or control), which are nested within farms. Farms serve as the primary blocking factor in the pairwise design, with each farm containing both a treatment field (with wildflower strip) and a control field (without wildflower strip).

Expected effect size and justification of sample size:

The expected effect size is Cohen's $d = 0.2$, representing a small but ecologically meaningful effect. This expectation is based on meta-analytic evidence suggesting variable but generally modest effects of wildflower strips on pollinator diversity (Albrecht et al., 2020; Scheper et al., 2015). A small effect size is anticipated because: (1) pollinator communities are influenced by multiple factors beyond local floral resources, including landscape context and regional species pools; (2) control fields may contain some floral resources from field margins or other semi-natural habitats; and (3) species richness is a relatively conservative metric that may be less sensitive than abundance measures.

Sample size justification:

Power analysis was conducted to determine the required number of farms and samples per farm

to detect Cohen's $d = 0.2$ with 80% power at $\alpha = 0.05$ (two-tailed). The analysis employed three complementary approaches: (1) a simple paired t-test calculation to establish a baseline, (2) design-effect adjusted calculations accounting for the nested structure, and (3) simulation-based power analysis using the planned mixed-effects model structure (see Appendix A for detailed calculations).

For a simple paired comparison, approximately 394 pairs would be required. However, the nested design with multiple samples per farm allows for more efficient estimation through the design effect approach. The design effect accounts for clustering within farms using the formula: $DE = 1 + (m - 1) \times ICC$, where m is the number of observations per farm and ICC is the intraclass correlation coefficient.

Assuming a mixed-effects model structure with farms as random effects, the required sample size depends on: - The number of farms (clusters) - The number of samples per treatment per farm - The intraclass correlation coefficient (ICC), representing the proportion of variance attributable to between-farm differences

Based on the power analysis calculations (see Appendix A), using typical ICC values for ecological field studies (0.1-0.3) and assuming 8-10 samples per treatment per farm per year (weekly sampling over a 12-16 week season), with two years of data collection:

- **Target sample size: 35-40 farms** with paired treatment and control fields
- **Samples per treatment per farm: 8-10 per year** (weekly during peak season)
- **Total sampling events: 560-800** ($35 \text{ farms} \times 2 \text{ treatments} \times 8-10 \text{ samples} \times 2 \text{ years}$)

This design provides adequate power ($>80\%$) to detect the expected small effect size while accounting for the hierarchical structure of the data. The pairwise design within farms controls for farm-level confounding factors, and the temporal replication (two years) allows for assessment of inter-annual variation.

Experimental design and sampling design:

The study employs a **pairwise experimental design** with **nested sampling**:

1. **Farm selection:** Farms are selected based on participation in wildflower strip programs, ensuring that each farm has at least one field with an established wildflower strip (treatment) and an adjacent or nearby control field without a wildflower strip.
2. **Pairwise matching:** Within each farm, treatment and control fields are matched to minimize

differences in soil type, field size, crop type, and landscape context. This pairing controls for farm-level and field-level confounding factors.

3. **Temporal replication:** Sampling occurs weekly during the peak pollinator season (typically May-September in the UK, approximately 12-16 weeks) over two consecutive years, providing both seasonal and inter-annual replication.
4. **Spatial replication:** Within each field, multiple pan trap stations are established (typically 3-5 stations per field, depending on field size), with traps placed at standardized distances from field edges and wildflower strips.
5. **Sampling protocol:** Pan traps (colored bowls: white, yellow, and blue) are placed at ground level, filled with soapy water, and left for 48 hours. Traps are collected, and pollinators are identified to species level where possible, or to morphospecies for taxonomic groups requiring specialist identification.

3.2 Variables

Dependent variable(s):

- **Pollinator species richness:** The number of unique pollinator species (or morphospecies) identified in each pan trap collection. This will be calculated per sampling event (trap collection) and may also be aggregated at the field level for some analyses.

Key predictors / exposures:

- **Treatment:** Binary variable indicating presence (1) or absence (0) of wildflower strip. This is the primary predictor of interest.
- **Year:** Categorical variable (Year 1, Year 2) to account for inter-annual variation.
- **Week:** Continuous or categorical variable representing sampling week within season (1-16) to account for seasonal phenology effects.

Covariates / confounders:

- **Farm size:** Continuous variable (hectares) to account for potential farm-scale effects.
- **Crop type:** Categorical variable (e.g., arable, mixed, pasture) to account for differences in agricultural management.
- **Field size:** Continuous variable (hectares) for the sampled field.

- **Landscape context:** Continuous variable representing the proportion of semi-natural habitat within a specified radius (e.g., 1 km) of the farm, derived from land-use maps.
- **Distance to nearest semi-natural habitat:** Continuous variable (meters) from the sampled field to the nearest hedgerow, woodland, or other semi-natural habitat.
- **Weather variables:** Mean temperature and precipitation during the sampling period, obtained from nearby weather stations.

Grouping / random effects:

- **Farm ID:** Random effect to account for clustering of samples within farms and to control for unmeasured farm-level characteristics.
- **Field ID:** Nested random effect (within farms) to account for repeated sampling of the same fields over time.
- **Trap station ID:** Nested random effect (within fields) if multiple trap stations are used per field, to account for spatial autocorrelation.

3.3 Data handling and quality control

Inclusion and exclusion criteria:

- **Inclusion:** All pan trap collections with complete identification data will be included in the primary analysis. Collections where traps were disturbed, damaged, or lost will be excluded.
- **Exclusion:** Samples collected during extreme weather events (e.g., heavy rain, strong winds) that may have affected trap effectiveness will be excluded. Farms where wildflower strips are less than one year old will be excluded to ensure establishment effects are minimal.
- **Taxonomic resolution:** Pollinators will be identified to species level where possible. For groups requiring specialist identification (e.g., some small bees, hoverflies), morphospecies will be used. Only clearly identifiable specimens will be included in species richness calculations.

Missing data approach:

- Missing trap collections (due to disturbance, weather, or logistical constraints) will be treated as missing at random, assuming that missingness is not systematically related to treatment condition or pollinator abundance.
- Missing covariate data (e.g., weather data) will be imputed using appropriate methods (e.g., mean imputation for continuous variables, or multiple imputation if substantial missingness is

present).

- Sensitivity analyses will be conducted excluding farms or time periods with substantial missing data to assess robustness of results.

Outliers and data validation:

- Species richness values will be examined for outliers using standard diagnostic plots (boxplots, Q-Q plots). Extreme values will be investigated to determine if they represent genuine biological variation or data errors.
- Outliers identified as data entry errors will be corrected or removed. Biological outliers (e.g., unusually high richness due to exceptional field conditions) will be retained but flagged for sensitivity analysis.
- Data validation will include: (1) cross-checking species identifications against reference collections and expert verification of difficult taxa; (2) verifying spatial coordinates and field assignments; (3) checking temporal consistency (sampling dates, year assignments).

Transformations or derived variables:

- Species richness is a count variable and may be modeled using Poisson or negative binomial distributions. If overdispersion is present, negative binomial models will be preferred.
- If using linear models, square root transformation of species richness may be considered to stabilize variance, though this will be evaluated through model diagnostics.
- Derived variables may include: (1) cumulative species richness per field per season; (2) species turnover (beta diversity) between treatment and control fields; (3) functional diversity metrics if sufficient trait data are available.

3.4 Statistical analysis strategy

Overall analytical approach:

The analysis will emphasize **confirmatory inference** for the primary objective (testing the effect of wildflower strips on species richness) while maintaining an **exploratory component** for secondary objectives and covariate effects. The primary analysis will be pre-specified and conducted as planned, with secondary analyses used to explore patterns and potential moderators.

Modelling philosophy:

The approach will be **explanatory**, focusing on understanding the causal or associative relationship between wildflower strips and pollinator species richness, while accounting for known confounding factors. Models will balance **parsimony** with biological realism, including relevant covariates and random effects to account for the nested structure of the data. Model selection will be guided by both statistical criteria (e.g., AIC, likelihood ratio tests) and biological knowledge.

Statistical framework:

A **frequentist framework** will be used for the primary analysis, with mixed-effects models estimated using maximum likelihood or restricted maximum likelihood. Confidence intervals and p-values will be reported for effect estimates. Effect sizes (Cohen's d) will be calculated and interpreted alongside statistical significance. Bayesian approaches may be considered for sensitivity analyses if computational resources permit.

3.5 Primary analysis

Statistical model(s) to be used:

The primary analysis will use a **linear mixed-effects model** (LMM) or **generalized linear mixed-effects model** (GLMM) depending on the distribution of the species richness data:

Model 1: Linear Mixed-Effects Model (if assumptions are met)

$$\text{Species_richness}_{ijkl} = \beta_0 + \beta_1 \text{Treatment}_{ij} + \beta_2 \text{Year}_k + \beta_3 \text{Week}_l +$$

$$\beta_4(\text{Treatment} \times \text{Year})_{ijk} + \beta_5(\text{Treatment} \times \text{Week})_{ijl} + \gamma_i + \varepsilon_{ijkl}$$

Where: - $\text{Species_richness}_{ijkl}$ is the species richness for sample l in field j (treatment/control) of farm i in year k - β_0 is the intercept - β_1 is the main effect of treatment (wildflower strip vs. control) - β_2 is the effect of year - β_3 is the effect of sampling week - β_4 and β_5 are interaction terms - $\gamma_i \sim N(0, \sigma_{\text{farm}}^2)$ is the random effect for farm i - $\varepsilon_{ijkl} \sim N(0, \sigma_{\text{residual}}^2)$ is the residual error

Model 2: Generalized Linear Mixed-Effects Model (if count distribution requires)

If species richness shows overdispersion or requires count modeling:

$$\text{Species_richness}_{ijkl} \sim \text{Negative Binomial}(\mu_{ijkl}, \theta)$$

$$\log(\mu_{ijkl}) = \beta_0 + \beta_1 \text{Treatment}_{ij} + \beta_2 \text{Year}_k + \beta_3 \text{Week}_l +$$

$$\beta_4(\text{Treatment} \times \text{Year})_{ijk} + \beta_5(\text{Treatment} \times \text{Week})_{ijl} + \gamma_i$$

Key predictors and adjustments:

- **Primary predictor:** Treatment (wildflower strip presence/absence), with adjustment for the pairwise design through farm-level random effects
- **Temporal adjustments:** Year and Week (and their interactions with treatment) to account for temporal variation
- **Covariate adjustments:** Landscape context, field size, and weather variables will be included as fixed effects if they show substantial correlation with the outcome or treatment
- **Random effects:** Farm ID (and potentially Field ID nested within Farm) to account for clustering and control for unmeasured farm-level confounders

Assumptions to be assessed:

1. **Normality of residuals:** Q-Q plots and Shapiro-Wilk tests for LMM; if violated, consider GLMM or transformations
2. **Homoscedasticity:** Residual plots against fitted values; if violated, consider variance structures or robust standard errors
3. **Independence:** Assessment of temporal autocorrelation (ACF/PACF plots) and spatial autocorrelation (if applicable)
4. **Linearity:** Scatterplots of residuals against continuous predictors
5. **Random effects distribution:** Q-Q plots of random effect predictions to assess normality assumption
6. **Overdispersion** (for count models): Assessment using dispersion parameters and residual diagnostics
7. **Influential observations:** Cook's distance and leverage diagnostics

Estimands or contrasts of interest:

1. **Primary estimand:** The difference in mean species richness between treatment and control fields, estimated as $\beta_{\text{Treatment}}$ in the model above. This will be reported as:
 - Raw difference (mean treatment - mean control)
 - Standardized effect size (Cohen's $d = (\beta_{\text{Treatment}} - \beta_{\text{Control}}) / \sigma_{\text{pooled}}$)
 - 95% confidence interval
 - p-value from two-tailed test ($H_0: \beta_{\text{Treatment}} = 0$)
2. **Secondary estimands:**
 - Treatment effect by year ($\beta_{\text{Treatment} \times \text{Year}}$ for each year)
 - Treatment effect by week (to assess seasonal variation)
 - Variance components (σ^2_{farm} , $\sigma^2_{\text{residual}}$) to quantify the proportion of variance attributable to farm-level differences

3.6 Secondary and sensitivity analyses

Alternative model specifications:

1. **Model without interactions:** Simplified model excluding Treatment \times Year and Treatment \times Week interactions to assess robustness of main effect estimate
2. **Model with additional covariates:** Inclusion of landscape context, field size, and weather variables as fixed effects
3. **Model with field-level random effects:** Addition of Field ID as a nested random effect (within Farm) if multiple fields per treatment are sampled
4. **Robust standard errors:** Use of robust (sandwich) standard errors to account for potential model misspecification
5. **Spatial correlation structures:** If spatial autocorrelation is detected, inclusion of spatial correlation structures (e.g., exponential, spherical) in the random effects

Subgroup or stratified analyses:

1. **By pollinator functional group:** Separate analyses for bees, hoverflies, and butterflies (if sample sizes permit) to assess whether effects vary by taxonomic group
2. **By landscape context:** Stratified analysis comparing effects in farms with high vs. low proportions of semi-natural habitat in the surrounding landscape

3. **By crop type:** Stratified analysis by dominant crop type (arable vs. mixed vs. pasture) to assess context-dependency
4. **By year:** Separate analyses for Year 1 and Year 2 to assess temporal consistency

Robustness checks or sensitivity analyses:

1. **Exclusion of outliers:** Re-analysis excluding identified outliers to assess their influence on effect estimates
2. **Exclusion of farms with missing data:** Re-analysis excluding farms with >20% missing sampling events
3. **Different transformation approaches:** Comparison of results using raw species richness, square-root transformed, and log-transformed outcomes
4. **Alternative effect size metrics:** Calculation of additional effect size measures (e.g., Hedges' g, odds ratio if using binary outcomes) for comparison
5. **Bootstrap confidence intervals:** Non-parametric bootstrap (1000 iterations) to generate confidence intervals, particularly if distributional assumptions are questionable
6. **Permutation tests:** Permutation-based p-values (1000 permutations) as a non-parametric alternative to assess statistical significance

Multiple testing considerations:

- The primary analysis (testing the main treatment effect) is a single pre-specified hypothesis test and does not require multiple testing correction.
- For secondary analyses involving multiple subgroup comparisons or interaction terms, the following approaches will be used:
 - **Bonferroni correction:** For exploratory subgroup analyses, p-values will be adjusted using Bonferroni correction based on the number of comparisons
 - **False Discovery Rate (FDR) control:** For exploratory analyses with many comparisons, Benjamini-Hochberg FDR correction will be applied
 - **Pre-specification:** Secondary analyses will be clearly distinguished from primary analyses, and any post-hoc exploratory findings will be clearly labeled as such

4 Outputs, Interpretation & Reporting

4.1 Planned outputs

Tables:

1. **Descriptive statistics table:** Mean, standard deviation, and range of species richness by treatment group, year, and overall, with sample sizes
2. **Primary model results table:**
 - Fixed effect estimates (coefficients) with standard errors, 95% confidence intervals, and p-values
 - Random effect variance components (farm-level variance, residual variance) with intraclass correlation coefficient
 - Model fit statistics (AIC, BIC, log-likelihood, conditional and marginal R^2)
3. **Effect size summary table:** Cohen's d with 95% confidence interval, interpreted using standard benchmarks (small: 0.2, medium: 0.5, large: 0.8)
4. **Sensitivity analysis summary table:** Comparison of treatment effect estimates across different model specifications and sensitivity analyses

Figures:

1. **Exploratory data visualization:**
 - Boxplots or violin plots of species richness by treatment group
 - Time series plots showing species richness across sampling weeks by treatment and year
 - Scatterplots of species richness against key covariates (landscape context, field size)
2. **Primary analysis results:**
 - Predicted species richness (with confidence intervals) by treatment group, adjusted for covariates
 - Effect plot showing the treatment effect estimate with 95% confidence interval
 - Interaction plots (if interactions are significant) showing treatment effects by year or week
3. **Model diagnostics:**
 - Q-Q plots of residuals and random effects
 - Residual plots (residuals vs. fitted values, residuals vs. predictors)
 - Leverage and influence diagnostics (Cook's distance plots)

- ACF/PACF plots for temporal autocorrelation assessment

4. **Sensitivity analysis visualization:**

- Forest plot comparing effect estimates across different model specifications
- Subgroup analysis plots showing treatment effects by landscape context or crop type (if conducted)

Summary statistics:

- Overall descriptive statistics (mean, median, SD, range, IQR) for species richness
- Summary by treatment group, year, and their combinations
- Variance decomposition: proportion of variance explained by fixed effects (marginal R^2) and by both fixed and random effects (conditional R^2)
- Intraclass correlation coefficient (ICC) quantifying the proportion of variance attributable to farm-level clustering

4.2 Interpretation principles

Emphasis on effect sizes and uncertainty:

- Statistical significance (p-values) will be reported but not emphasized as the sole criterion for interpretation. The primary focus will be on the **magnitude and direction of the effect** (effect size) and the **precision of the estimate** (confidence intervals).
- Cohen’s d will be interpreted using standard benchmarks (Cohen, 1988): small (0.2), medium (0.5), large (0.8), with recognition that even small effects can be ecologically meaningful in conservation contexts.
- Confidence intervals will be interpreted to assess both the magnitude and precision of the effect. Narrow intervals indicate precise estimates; intervals that exclude zero indicate statistical significance at the chosen level.
- The expected effect size (Cohen’s $d = 0.2$) represents a small effect, and the study is powered to detect this effect. If the observed effect is smaller than expected, this will be discussed in the context of study limitations and potential reasons (e.g., high baseline diversity in control fields, insufficient establishment time for wildflower strips).

Scientific or practical relevance:

- Results will be interpreted in the context of existing literature, particularly the meta-analyses

by Albrecht et al. (2020) and Scheper et al. (2015), to assess whether findings are consistent with previous research.

- The practical relevance of the effect size will be discussed in terms of conservation implications: even small increases in pollinator species richness may contribute to ecosystem resilience and pollination services, particularly when implemented at landscape scales.
- If the effect is non-significant or smaller than expected, this will be interpreted as evidence that wildflower strips may have context-dependent effectiveness, requiring consideration of landscape context, farm management, and other local factors.

Limitations and caveats:

- **Causal interpretation:** The pairwise design within farms helps control for farm-level confounders, but causal inference is limited by the observational nature of the study. Farms self-select into wildflower strip programs, and treatment and control fields may differ in unmeasured ways. Results should be interpreted as associations rather than causal effects, though the pairwise design strengthens inference compared to completely observational studies.
- **Generalizability:** Results are specific to UK farms participating in wildflower strip programs and may not generalize to other regions, farm types, or conservation contexts.
- **Temporal scope:** Two years of data may be insufficient to capture long-term effects of wildflower strips, which may require several years to establish and reach full effectiveness.
- **Taxonomic resolution:** Use of morphospecies for some taxa may introduce measurement error, though this is expected to be non-differential with respect to treatment.
- **Sampling method:** Pan traps may underrepresent certain pollinator groups (e.g., butterflies, which are better sampled by transect walks). Results should be interpreted as effects on pan-trap-captured pollinator communities.
- **Missing data:** If substantial missing data occur, results may be sensitive to assumptions about missingness mechanisms, despite planned sensitivity analyses.

5 Reproducibility & Amendments

Software and versions:

- **R version:** 4.3.0 or later (or current version at time of analysis)
- **Key R packages:**

- `lme4` (version 1.1-34 or later) for mixed-effects models
- `glmmTMB` (version 1.1.7 or later) for generalized linear mixed-effects models if needed
- `emmeans` (version 1.8.8 or later) for estimated marginal means and effect size calculations
- `performance` (version 0.10.4 or later) for model diagnostics and R^2 calculations
- `ggplot2` (version 3.4.2 or later) for data visualization
- `dplyr` (version 1.1.2 or later) for data manipulation
- `effectsize` (version 0.8.3 or later) for effect size calculations
- **Documentation:** Quarto (version 1.3 or later) for generating the analysis report
- All package versions will be recorded using `renv` or `sessionInfo()` output

Reproducibility approach:

- **Analysis scripts:** All analyses will be conducted using R scripts organized in a clear directory structure:
 - `01_data_cleaning.R`: Data import, cleaning, and preparation
 - `02_exploratory_analysis.R`: Exploratory data analysis and visualization
 - `03_primary_analysis.R`: Primary statistical models and effect size calculations
 - `04_secondary_analysis.R`: Secondary and sensitivity analyses
 - `05_figures_tables.R`: Generation of publication-ready figures and tables
- **Quarto notebook:** A Quarto document (`analysis_report.qmd`) will integrate all analyses with narrative text, generating a complete analysis report
- **Version control:** All code and data will be managed using Git, with commits documenting major analysis steps. A GitHub or GitLab repository will be maintained with clear commit messages.
- **Data management:** Raw data will be stored in `data/raw/`, cleaned data in `data/processed/`, with clear documentation of all data transformations
- **Seed setting:** Random number generators will use set seeds (e.g., `set.seed(12345)`) for any stochastic processes (bootstrap, permutation tests) to ensure reproducibility
- **Code documentation:** All functions and complex analyses will include inline comments explaining the rationale and approach

Criteria for deviations from this plan and how they will be documented:

- **Minor deviations** (e.g., addition of covariates discovered during exploratory analysis, use of alternative transformation): These will be documented in the analysis script comments and

reported in the methods section of any resulting publication. The primary analysis will be conducted as planned, with deviations reported as sensitivity analyses.

- **Major deviations** (e.g., change in primary model structure, modification of sample size, change in primary outcome): These require justification and will be:
 1. Documented in a “Deviations from Analysis Plan” section in the analysis report
 2. Clearly explained with rationale (e.g., violation of model assumptions requiring alternative approach)
 3. Reported alongside the original planned analysis where possible
 4. Discussed with co-authors or statistical consultants before implementation
- **Post-hoc analyses:** Any analyses not pre-specified in this plan will be clearly labeled as “exploratory” or “post-hoc” in all outputs and publications
- **Amendment log:** A dated log of any amendments to this plan will be maintained, documenting the date, nature of change, and rationale

5.1 Optional sign-off (as required)

- Researcher:
- Statistical consultant:
- Supervisor:
- Date:

6 Resources

Study-specific references:

Albrecht, M., Kleijn, D., Williams, N.M., Tschumi, M., Blaauw, B.R., Bommarco, R., Campbell, A.J., Dainese, M., Drummond, F.A., Entling, M.H., Ganser, D., Arjen de Groot, G., Goulson, D., Grab, H., Hamilton, H., Herzog, F., Isaacs, R., Jacot, K., Jeanmeret, P., Jonsson, M., Knop, E., Kremen, C., Landis, D.A., Loeb, G.M., Marini, L., McKerchar, M., Morandin, L., Pfister, S.C., Potts, S.G., Rundlöf, M., Sardiñas, H., Sciligo, A., Thies, C., Tscharntke, T., Venturini, E., Veromann, E., Vollhardt, I.M.G., Wäckers, F., Ward, K., Westbury, D.B., Wilby, A., Woltz, M., Wratten, S., Sutter, L., 2020. The effectiveness of flower strips and hedgerows on pest control, pollination services and crop yield: a quantitative synthesis. *Ecology Letters* 23, 1488–1498.

Scheper, J., Bommarco, R., Holzschuh, A., Potts, S.G., Riedinger, V., Roberts, S.P.M., Rundlöf, M., Smith, H.G., Steffan-Dewenter, I., Wickens, J.B., Wickens, V.J., Kleijn, D., 2015. Local and landscape-level floral resources explain effects of wildflower strips on wild bees across four European countries. *Journal of Applied Ecology* 52, 1165–1175.

Methodological references:

Cressman, K.A., Sharp, J.L., 2022. Crafting statistical analysis plans: A cross-discipline approach. *Stat* 11, e528.

Gamble, C., Krishan, A., Stocken, D., Lewis, S., Juszczak, E., Doré, C., Williamson, P.R., Altman, D.G., Montgomery, A., Lim, P., Berlin, J., Senn, S., Day, S., Barbachano, Y., Loder, E., 2017. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. *JAMA* 318, 2337–2343.

Simpson, S.H., 2015. Creating a Data Analysis Plan: What to Consider When Choosing Statistics for a Study. *Can J Hosp Pharm* 68, 311–317.

Stevens, G., Dolley, S., Mogg, R., Connor, J.T., 2023. A template for the authoring of statistical analysis plans. *Contemp Clin Trials Commun* 34, 101100.

Yuan, I., Topjian, A.A., Kurth, C.D., Kirschen, M.P., Ward, C.G., Zhang, B., Mensinger, J.L., 2019. Guide to the statistical analysis plan. *Pediatric Anaesthesia* 29, 237–242.

7 Appendix A: Power Analysis Calculations

This appendix provides the detailed R code and calculations for the power analysis used to justify the sample size. The analysis employs three complementary approaches: (1) simple paired t-test calculation, (2) design-effect adjusted calculations for nested designs, and (3) simulation-based power analysis using the planned mixed-effects model structure.

7.1 A.1 Simple Paired t-Test Power Calculation

```
# Load required packages
library(pwr)

# Simple paired t-test power calculation
# Cohen's d = 0.2, power = 0.8, alpha = 0.05 (two-tailed)
```

```
power_result <- pwr.t.test(
  d = 0.2,          # Effect size (Cohen's d)
  power = 0.8,      # Desired power
  sig.level = 0.05, # Alpha level
  type = "paired",  # Paired design
  alternative = "two.sided"
)
```

Display results

```
cat("Simple paired t-test power analysis:\n")
```

Simple paired t-test power analysis:

```
cat("Required number of pairs (farms):", ceiling(power_result$n), "\n")
```

Required number of pairs (farms): 199

```
cat("Actual power with n =", ceiling(power_result$n), ":",
    round(power_result$power, 3), "\n\n")
```

Actual power with n = 199 : 0.8

7.2 A.2 Nested Design Power Analysis Using Design Effect

```
# Power analysis for nested design using design effect
# The design effect accounts for clustering within farms

# Function to calculate design effect and effective sample size
# Design effect = 1 + (m - 1) * ICC
# where m = number of observations per cluster
# ICC = intraclass correlation coefficient

calculate_power_nested <- function(
  d,          # Cohen's d
  n_farms,    # Number of farms (clusters)
```

```

n_samples_per_treatment_per_farm, # Samples per treatment per farm
icc, # Intraclass correlation coefficient
alpha = 0.05,
power_target = 0.8
) {
  # Total observations per farm (treatment + control)
  m <- n_samples_per_treatment_per_farm * 2

  # Design effect
  deff <- 1 + (m - 1) * icc

  # Effective sample size (adjusted for clustering)
  n_effective <- n_farms / deff

  # Power calculation for effective sample size
  power_calc <- pwr.t.test(
    d = d,
    n = n_effective,
    sig.level = alpha,
    type = "paired",
    alternative = "two.sided"
  )

  return(list(
    n_farms = n_farms,
    n_samples_per_treatment = n_samples_per_treatment_per_farm,
    m = m,
    icc = icc,
    design_effect = deff,
    n_effective = n_effective,
    power = power_calc$power
  ))

```

```

}

# Test different scenarios
scenarios <- expand.grid(
  n_farms = seq(20, 50, by = 5),
  n_samples = c(8, 10, 12),
  icc = c(0.1, 0.2, 0.3)
)

# Calculate power for each scenario
results <- apply(scenarios, 1, function(x) {
  calc <- calculate_power_nested(
    d = 0.2,
    n_farms = x[1],
    n_samples_per_treatment_per_farm = x[2],
    icc = x[3]
  )
  return(data.frame(
    n_farms = calc$n_farms,
    n_samples_per_treatment = calc$n_samples_per_treatment,
    icc = calc$icc,
    design_effect = round(calc$design_effect, 2),
    n_effective = round(calc$n_effective, 1),
    power = round(calc$power, 3)
  ))
})

power_table <- do.call(rbind, results)

# Filter for scenarios with power >= 0.8
adequate_power <- power_table[power_table$power >= 0.8, ]

```

```
cat("Power analysis results for nested design (Cohen's d = 0.2,  = 0.05):\n")
```

Power analysis results for nested design (Cohen's d = 0.2, = 0.05):

```
cat("Scenarios achieving 80% power:\n\n")
```

Scenarios achieving 80% power:

```
print(adequate_power[order(adequate_power$n_farms, adequate_power$n_samples_per_treatment), ],
      row.names = FALSE)
```

```
[1] n_farms          n_samples_per_treatment icc
[4] design_effect    n_effective             power
<0 rows> (or 0-length row.names)
```

```
# Recommended scenario
```

```
cat("\n--- Recommended Design ---\n")
```

--- Recommended Design ---

```
recommended <- adequate_power[
  adequate_power$n_farms >= 35 &
  adequate_power$n_farms <= 40 &
  adequate_power$n_samples_per_treatment >= 8 &
  adequate_power$n_samples_per_treatment <= 10 &
  adequate_power$icc == 0.2,
]

if(nrow(recommended) > 0) {
  rec <- recommended[1, ]
  cat("Farms:", rec$n_farms, "\n")
  cat("Samples per treatment per farm:", rec$n_samples_per_treatment, "\n")
  cat("ICC:", rec$icc, "\n")
  cat("Design effect:", rec$design_effect, "\n")
  cat("Effective sample size:", rec$n_effective, "\n")
}
```

```

cat("Power:", rec$power, "\n")
cat("\nTotal sampling events (2 years):",
    rec$n_farms * 2 * rec$n_samples_per_treatment * 2, "\n")
}

```

7.3 A.3 Simulation-Based Power Analysis for Mixed-Effects Model

```

# Simulation-based power analysis for mixed-effects model
# This provides a more accurate assessment for the actual model structure

library(lme4)

# Try to load lmerTest for p-values, but fall back to manual calculation if not available
has_lmerTest <- requireNamespace("lmerTest", quietly = TRUE)
if (has_lmerTest) {
  library(lmerTest)
}

# Set seed for reproducibility
set.seed(12345)

# Simulation parameters
n_sims <- 1000 # Number of simulations (reduce for faster computation)
d_true <- 0.2 # True effect size (Cohen's d)
alpha <- 0.05

# Function to simulate data and test hypothesis
simulate_and_test <- function(
  n_farms,
  n_samples_per_treatment_per_farm_per_year,
  n_years = 2,
  icc = 0.2,

```

```

d = d_true,
sigma_residual = 1
) {
  # Calculate variance components
  # ICC = sigma_farm^2 / (sigma_farm^2 + sigma_residual^2)
  # Therefore: sigma_farm^2 = ICC * sigma_residual^2 / (1 - ICC)
  sigma_farm <- sqrt(icc * sigma_residual^2 / (1 - icc))

  # Treatment effect in standard deviation units
  treatment_effect <- d * sigma_residual

  # Generate data
  farm_ids <- rep(1:n_farms, each = n_samples_per_treatment_per_farm_per_year * 2 * n_years)
  treatment <- rep(c(0, 1), times = n_farms * n_samples_per_treatment_per_farm_per_year * n_years)
  year <- rep(rep(1:n_years, each = n_samples_per_treatment_per_farm_per_year * 2), times = n_farms * n_years)

  # Farm-level random effects
  farm_effects <- rnorm(n_farms, mean = 0, sd = sigma_farm)
  farm_effects_expanded <- rep(farm_effects, each = n_samples_per_treatment_per_farm_per_year * 2)

  # Residual error
  residual <- rnorm(length(farm_ids), mean = 0, sd = sigma_residual)

  # Generate species richness (count variable, but modeled as continuous for simplicity)
  species_richness <- 10 + treatment_effect * treatment + farm_effects_expanded + residual
  species_richness <- pmax(0, round(species_richness)) # Ensure non-negative integers

  # Create data frame
  data <- data.frame(
    farm_id = factor(farm_ids),
    treatment = factor(treatment),
    year = factor(year),

```

```

    species_richness = species_richness
  )

# Fit mixed-effects model
result <- tryCatch({
  # Fit model - use lmerTest::lmer if available for p-values, otherwise lme4::lmer
  if (has_lmerTest) {
    model <- lmerTest::lmer(species_richness ~ treatment + year + (1 | farm_id),
                          data = data)
  } else {
    model <- lme4::lmer(species_richness ~ treatment + year + (1 | farm_id),
                      data = data)
  }

  # Check for critical convergence failures
  if (!is.null(model@optinfo$conv$lme4$messages)) {
    conv_msgs <- model@optinfo$conv$lme4$messages
    if (any(grepl("failed to converge|singular", conv_msgs, ignore.case = TRUE))) {
      return(NA)
    }
  }

  # Extract coefficient summary
  coef_summary <- summary(model)$coefficients

  # Find treatment coefficient
  # When treatment is factor with levels 0,1, coefficient is "treatment1"
  coef_names <- rownames(coef_summary)
  treatment_coef <- coef_names[grepl("^treatment", coef_names, ignore.case = TRUE)]

  if (length(treatment_coef) == 0) {
    return(NA)
  }

```



```

}

# Extract p-value
# lmerTest provides "Pr(>|t|)", lme4 doesn't provide p-values
if (has_lmerTest && "Pr(>|t|)" %in% colnames(coef_summary)) {
  p_value <- coef_summary[treatment_coef[1], "Pr(>|t|)"]
} else {
  # Fallback: calculate p-value from t-statistic (using large-sample approximation)
  # This is less accurate but works when lmerTest is not available
  t_stat <- coef_summary[treatment_coef[1], "t value"]
  if (is.na(t_stat) || !is.finite(t_stat)) {
    return(NA)
  }
  # Use t-distribution with large df (conservative)
  p_value <- 2 * (1 - pt(abs(t_stat), df = 1000))
}

# Check if p-value is valid
if (is.na(p_value) || !is.finite(p_value)) {
  return(NA)
}

return(p_value < alpha)
}, error = function(e) {
  return(NA)
})

return(result)
}

# Test recommended design

```

```
cat("Simulation-based power analysis for recommended design:\n")
```

Simulation-based power analysis for recommended design:

```
cat("(This may take a few minutes...)\n\n")
```

(This may take a few minutes...)

```
n_farms_rec <- 35
```

```
n_samples_rec <- 10
```

```
icc_rec <- 0.2
```

```
# First, test a single simulation to debug
```

```
cat("Testing a single simulation...\n")
```

Testing a single simulation...

```
test_result <- simulate_and_test(  
  n_farms = n_farms_rec,  
  n_samples_per_treatment_per_farm_per_year = n_samples_rec,  
  n_years = 2,  
  icc = icc_rec  
)  
cat("Single test result:", test_result, "\n\n")
```

Single test result: TRUE

```
# Run simulations (using fewer for demonstration)
```

```
n_sims_demo <- 200 # Use 200 for faster demonstration; increase to 1000 for final analysis
```

```
sim_results <- replicate(n_sims_demo,  
  simulate_and_test(  
    n_farms = n_farms_rec,  
    n_samples_per_treatment_per_farm_per_year = n_samples_rec,  
    n_years = 2,  
    icc = icc_rec
```

```

)
)

# Calculate power (proportion of significant results)
# Count valid (non-NA) results
n_valid <- sum(!is.na(sim_results))
n_significant <- sum(sim_results, na.rm = TRUE)

# Calculate power
power_sim <- ifelse(n_valid > 0, n_significant / n_valid, NA)

cat("Recommended design:\n")

```

Recommended design:

```
cat("  Farms:", n_farms_rec, "\n")
```

Farms: 35

```
cat("  Samples per treatment per farm per year:", n_samples_rec, "\n")
```

Samples per treatment per farm per year: 10

```
cat("  Years:", 2, "\n")
```

Years: 2

```
cat("  ICC:", icc_rec, "\n")
```

ICC: 0.2

```
cat("  Valid simulations:", n_valid, "out of", n_sims_demo, "\n")
```

Valid simulations: 200 out of 200

```
if (n_valid > 0) {
  cat("  Simulated power (n =", n_valid, "valid simulations):",
      round(power_sim, 3), "\n")
}
```

```

# Calculate confidence interval only if we have valid results
if (n_significant >= 0 && n_significant <= n_valid && n_valid > 0) {
  ci_result <- binom.test(n_significant, n_valid, conf.level = 0.95)
  cat(" 95% CI for power: [",
      round(ci_result$conf.int[1], 3), ", ",
      round(ci_result$conf.int[2], 3), "]\n")
} else {
  cat(" Note: Unable to calculate confidence interval due to insufficient valid results\n")
}
} else {
  cat(" Warning: All simulations failed. Check model specification.\n")
}

```

Simulated power (n = 200 valid simulations): 0.955

95% CI for power: [0.916 , 0.979]