Analysis Plan: The Effect of Birdsong on Self-Reported Mental Health

By ANTHONY HIGNEY, CLAIRE BUCHAN, SIMON BUTLER, MELISSA MARSELLE, ELEANOR RATCLIFFE, KONRAD UEBEL, AND NICK HANLEY.*

This is a pre-analysis plan for a study examining the causal impact of birdsong on self-reported well-being. It uses survey data combined with spacial data on acoustic indices. We intend to study the effect of cumulative exposure to birdsong rather than an immediate hedonic effect. The sections below detail the hypotheses, research design, data, and threats to identification and limitations.

I. Research Question

What is the impact of changes in birdsong, in a person's local area, on self-reported mental well-being in the UK?

A. Hypotheses

Our hypotheses are specified for one-tailed tests. This, along with the variables in the hypotheses are discussed in sections below.

1) **H1**: An increase in the acoustic complexity index increases self-reported mental well-being
   H0: An increase in the acoustic complexity index does not increase self-reported mental well-being

2) **H1**: An increase in bio acoustic index increases self-reported mental well-being
   H0: An increase in bio acoustic index does not increase self-reported mental well-being

3) **H1**: There is a non-monotonic relationship between the acoustic entropy score and self-reported mental well-being. Initially, an increase in acoustic entropy increases mental-well being, but further increases reduce well-being. In a regression of well-being on acoustic entropy the linear term on entropy will be positive and the squared term negative, and they will be jointly significant at the 0.05 significance level

* Anthony Higney: University of Glasgow, Anthony.Higney@glasgow.ac.uk, corresponding author. Claire Buchan: University of East Anglia, c.buchan@uea.ac.uk. Simon Butler: University of East Anglia, simon.j.butler@uea.ac.uk. Melissa Marselle: University of Surrey, m.marselle@surrey.ac.uk. Eleanor Ratcliffe: University of Surrey, eleanor.ratcliffe@surrey.ac.uk. Konrad Uebel: University of Surrey, k.uebel@surrey.ac.uk. Nick Hanley: University of Glasgow, Nicholas.Hanley@glasgow.ac.uk.

H0: An increase in acoustic entropy score does not decrease self-reported mental well-being

II. Design

A. *Intuition*

Our intent is to provide an estimate of a causal effect of cumulative exposure to birdsong on self-reported well-being. We will combine survey data with well-being questions and spacial data on the prevalence of birdsong in different areas of the UK. Below is a causal diagram in the form of a directed acyclic graph (DAG). This shows our assumed model of the path through which birdsong affects mental well-being (Y). We assume that the birdsong actually heard by the people in our sample is a product of the characteristics of the area (urban/rural, near roads/isolated etc) and the characteristics of the local bird population (which species and number of individual birds). These bird characteristics are themselves the result of local area characteristics (some species prefer wooded areas for example) and the season (generally, more birds in spring). The area characteristics are in turn affected by the season (e.g. deciduous trees lose their leaves in winter). All of these variables may affect mental health, Y.

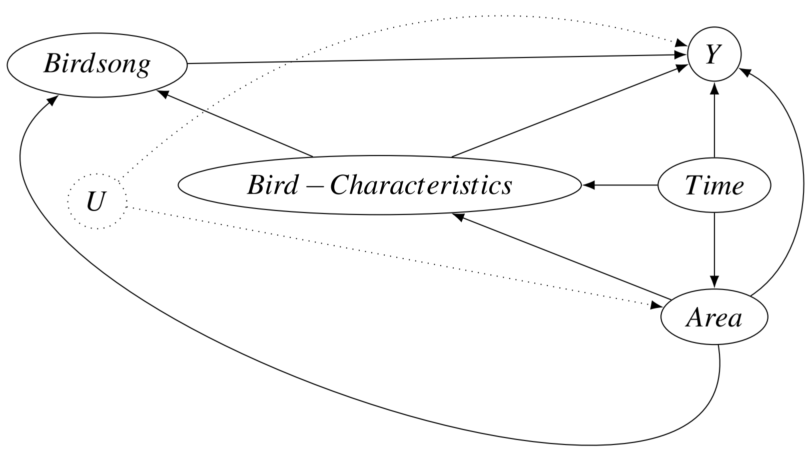

FIGURE 1. DAG MODEL OF THE EFFECT OF BIRDSONG ON MENTAL WELL-BEING

Given the above, simply regressing mental well-being on a measure of birdsong will not give a causal estimate of the effect. This is because there are confounding factors which cause both changes in mental well-being and changes in the birdsong heard, such as the season and the characteristics of the area. Furthermore, there are unobserved variables, U in the diagram, which will affect mental well-being and, indirectly, the birdsong heard by an individual. For example, people with higher incomes may choose to live in areas with higher amounts of birdsong, but their income may also lead to higher mental well-being. This would lead us to overestimate any effect of birdsong. In contrast, noise sensitive people may also choose to live in areas with less traffic noise and therefore more birdsong, but this noise sensitivity may affect their well-being negatively, leading us to underestimate the effect of birdsong.

To mitigate these issues, we will use a difference-in-differences design. We assume the outcome Y_{it}, for individual i at time t, is made up of individual and area time invariant factors, as well as time varying factors. We will observe the change in outcome for each individual in our sample between winter and spring, and the birdsong in their area in winter and spring. We call the outcome $Y_{it}(D_{it})$, and the level of birdsong D_{it}. By taking the first difference we remove time invariant factors. This is given by $\Delta Y_i = Y_{i2}(D_{1i}, \delta_i) - Y_{i1}(D_{1i})$, for a unit experiencing birdsong levels D_{1i} in winter and $D_{1i} + \delta_i$ in spring. Next we must remove all non-birdsong time-varying factors. This is more difficult because we do not assign different units to receive different treatment levels. This means other time varying factors may confound the effect of birdsong. Instead we must impute a counterfactual trend that would exist without the change in birdsong, $Y_{i2}(D_{1i}, \delta i) - \hat{Y}_{i2}(D_{1i}, 0)$. This is our second difference, hence difference-in-differences. To estimate $\hat{Y}_{i2}(D_{1i}, 0)$, given we have no comparable untreated units as a control group, we use ”quasi-stayers” (de Chaisemartin, D'Haultfœuille and Vazquez-Bare, 2024). These are units that experience small changes in birdsong between winter and spring. Furthermore, we will condition on observable factors and the level of birdsong in winter. So units are being compared firstly with their previous winter levels of well-being, and secondly with the trend in well-being for units that had similar levels of birdsong in period 1, similar levels of income, education etc, but different levels of birdsong in period 2.

The crucial assumption is that, conditional on our covariates, units that experience the same level of treatment in period one, **on average** would have had the same changes in outcome **but for the change in treatment intensity**. This is known as a parallel trends assumption. A simplified illustration of this is in figure 2.

Formally:

$$E[\Delta Y(d) \mid \mathbf{x}, D_1 = d, D_2] = E[\Delta Y(d) \mid \mathbf{x}, D_1 = d], \; \forall d \in D_1. \tag{1}$$

If units sort themselves into areas, such that units that respond the most to birdsong sort into areas that experience the largest changes in birdsong, then this is not a threat to our identification strategy. This is because the counterfactual for those units is no large increase in birdsong, so it does not matter if they respond more favourably or not once they receive the treatment. However, our estimand (equation 3) is analogous to

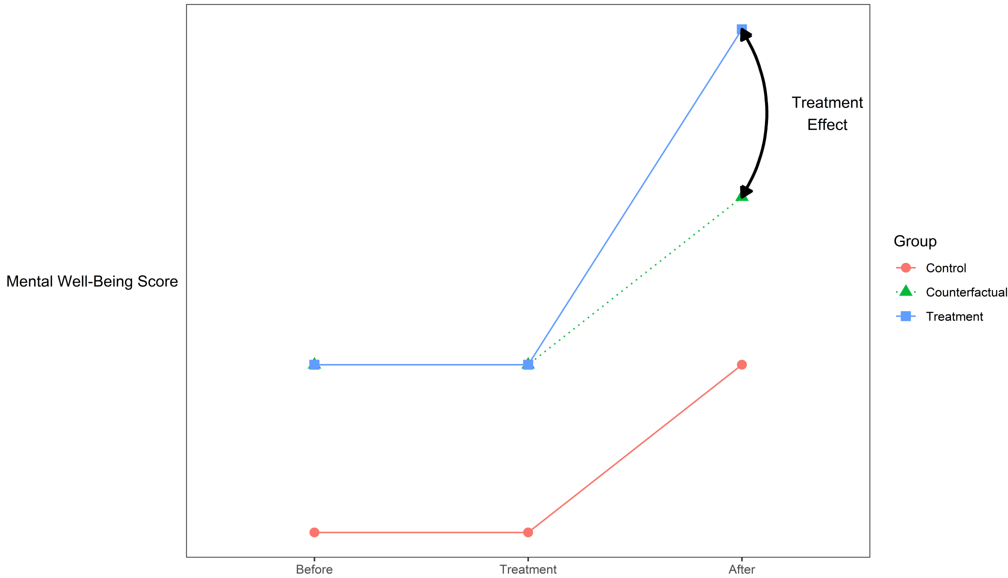

FIGURE 2. DIFFERENCE IN DIFFERENCES ILLUSTRATION OF PARALLEL TRENDS

the average effect of treatment on the treated (ATT), therefore our estimate cannot say anything about treatment on the *untreated*. This matters for the generalizability of our results but not the internal validity.

However, there are several threats to causal identification. The main threat is that other time-varying factors, correlated with both mental well-being and birdsong, are not accounted for by the imputed trend. This could be, for example, when two individuals both live in an area with otherwise similar characteristics, and individually similar on observables, but one lives much closer to green-space. As winter turns to spring, the flowers bloom and the tree leaves regrow, their mental well-being would be boosted more, even without any change in birdsong, than the similar person who lives slightly further away from the green-space. We aim to mitigate this threat by including area level green-space controls: garden access and urban/rural/suburban controls.

A second threat comes from if people, due to unobserved differences, systematically and directly improve the local birdsong, rather than sort into areas where there is more birdsong. This could be, for example, setting up many local birdfeeders in a way that greatly alters our birdsong variable, or they successfully campaign to improve local bird habitats. On our diagram it would appear as an additional arrow from U to the bird characteristics.

To mitigate this, we include questions in our survey that indicate level of knowledge in birds and whether respondents take part in bird watching. We will report regressions with and without such ”bird fans”. An additional limitation is that our birdsong variables

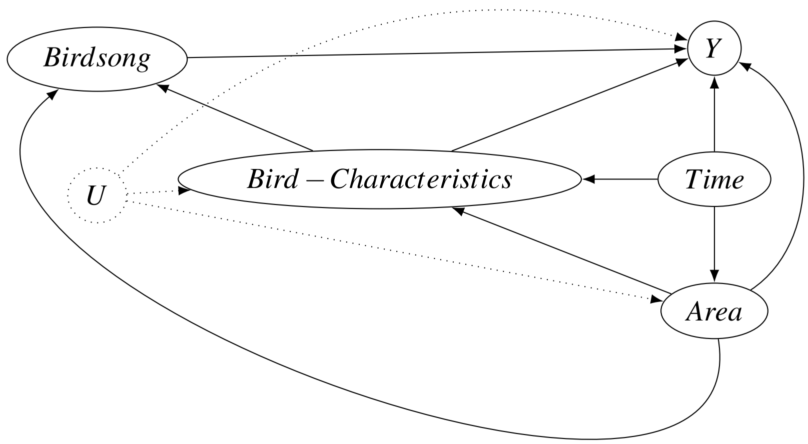

FIGURE 3. DAG MODEL OF THE EFFECT OF BIRDSONG ON MENTAL WELL-BEING WITH ADDITIONAL CONFOUNDER

are derived from bird characteristics data. therefore it is difficult to separate out the well-being effect of seeing birds, from that of hearing them. So our estimate may be biased upwards, or it may be better to think of it as the total well-being effect of birds. Again, we will include regressions with and without people who have extensive bird knowledge and report bird watching , as we expect the majority of the sight effect would be concentrated in these people.

B. Formal Design

RECRUITMENT

We will carry out a panel survey in two waves. This will sample the same individuals in both winter and spring. The first wave will be carried out in February and early March. The spring wave will be at the height of the ”Dawn Chorus” when morning birdsong is at its peak in most parts of the UK. We expect high rates of attrition so 3,500 will be sampled in the first wave, so that a minimum of 700 can be sampled in the second wave (with 1000 expected). The survey will be administered by Qualtrics LLC.

Each respondent will supply a UK postcode. This allows us to match the area they live with the local soundscapes variables (described in section III).

Balance Tests

Due to the large expected attrition rate (around 70 percent) expected in the online panel, we will carry out a joint covariate balance test (Hansen and Bowers, 2008) comparing attritioned respondents to those in the second wave and report results.

III. Data

Our design combines two sources of data: a panel survey for the outcome variables and control variables, combined with acoustic indices from simulated soundscapes calculated from UK-Breeding Bird Survey (BBS) data for the treatment variables.

A. Variables

Outcomes Variable

WHO-5 Index: This is a set of five questions developed by the World Health Organisation (WHO) to assess self-reported mental well-being. Responses are on a 6-point Likert scale ranging from ”All of the Time” to ”Not at All”. Participants are asked to rate, based on the last two weeks, the following five questions using the scale:

> I have felt cheerful and in good spirits.
>
> I have felt calm and relaxed.
>
> I have felt active and vigorous.
>
> I woke up feeling fresh and rested.
>
> My daily life has been filled with things that interest me.

The responses are given a score, where ”All of the Time” = 5, and ”Not at All” = 0. These are summed into an index score with the values 0-25. This is the WHO-5 Index score for that respondent. The values are normally then multiplied by four but this step is superfluous and we do not do that.

The WHO-5 index has high validity, sensitivity and specificity as well as being only five quick and non-invasive questions (Topp et al., 2015). Therefore, we judge it is a suitable metric for gauging self-reported well-being through a survey.

Treatment Variables

We use three main treatment variables. All three are modelled acoustic indices. These use UK-Breeding Bird Survey (BBS) and Bird Atlas 2007–11 survey data combined with sound files from Xeno Canto www.xeno-canto.org.

The BBS has been running since 1994. The UK is divided into 1km squares, which are sampled at least twice in spring. All bird visual and auditory defections are recorded and the species is noted. This gives a detailed survey of both species presence and numbers.

The Bird Atlas was a similar survey, except 4km^2 units were surveyed, and the survey was carried out in both winter and spring for the years 2007-11.

This survey data was combined with sound recordings of the species uploaded onto the xeno canto site. This is done following the method in Morrison et al. (2021). The xeno canto site has around 900,000 recordings of bird vocalisations from all over the world. Sound files were created based on the survey species and numbers, using the recordings of those species. Distance was simulated by sampling the volume from a uniform distribution for each individual. This process was repeated 100 times for each square. From these, three acoustic metrics were measured: Acoustic Complexity Index (ACI), Bioacoustic Index (Bio), and Acoustic entropy (H). These are further described in Abrahams et al. (2023), but we include a short description below. See Ratcliffe, Gatersleben and Sowden (2016) and Ratcliffe (2021) for more on why we would expect such variables to positively affect our outcome variable.

- **Acoustic Complexity Index**: This is a measure of the variation in acoustic intensity. Birdsong will tend to produce higher values of the ACI, whereas constant noise produces low values. We expect there to be a positive relationship between this variable and mental well-being, but perhaps with diminishing marginal effects at higher values.

- **Bioacoustic Index**: High values of this index occur when there is a large disparity in volume between the quietest and loudest parts of a recording. Therefore increases in this index tend to be associated with increases in bird vocalizations. We expect there to be a positive relationship between this variable and mental well-being, but perhaps with diminishing marginal effects at higher values.

- **Acoustic Entropy**: High values can either mean extremely noisy soundscapes or very quiet soundscapes. Low values mean that noise is concentrated in a small band of frequencies. We expect that mental well-being will initially increase as acoustic entropy increases, as more birdsong is heard from more individuals, but this relationship will reverse at higher values of entropy. This is because some birdsong is preferred to completely silent ones.

Control Variables

In every test we will include in a specification with only area, season, and treatment variables, as well as a specification that includes a number of control variables, both in the interests of more efficient estimation and to test the conditional parallel trends assumption. These variables are: income. age, education, self-reported visits to local green spaces, access to a garden, self-reported visits to other greenspaces, access to private transport, noise-sensitivity, and is the area urban, rural, a village or suburban, as well as which country/region of the UK they live in.

IV. Estimation

We will report the following regression as a benchmark:

$$Y_{it} = \alpha_i + \lambda_t + \theta D_{it} + \mathbf{x}_i\beta + \varepsilon_{it} \tag{2}$$

Where Y_{it} is the WHO5 score, α_i is the time-invariant individual fixed effect, λ_t is the time-varying fixed-effect, which equals 1 in spring and 0 in winter, D_{it} is the soundscape metric (i.e. the treatment variable), and $\mathbf{x_i}$ a vector of covariates such as income and education (see section III).

However, we believe this will be a biased estimate of the effect of birdsong due to differential sorting into different areas, as explained above. Therefore, our main estimate will use a difference-in-difference design. Given we have continuous treatment variables, and no untreated units, as well as no units who do not change treatment levels, we cannot estimate with two-way fixed effects. Two-way fixed effects is not robust to treatment effect heterogeneity, and we may not be able to identify the effect of treatment (Callaway, Goodman-Bacon and Sant'Anna, 2024). Instead will will use the method in de Chaisemartin, D'Haultfœuille and Vazquez-Bare (2024). This allows for treatment heterogeneity, a continuous treatment variable, and, importantly, no untreated units, or units that do not change treatment.

This method relies on the existence of "quasi-stayers". That is, units who change treatment an arbitrary, small amount. Within our sample there will be some areas with far less change in birdsong that others and this allows us to estimate the counter-factual trend.

Our main estimand is the weighted average marginal effect of treatment. This is average effect of moving from the treatment in period one to the treatment in period two, scaled by the average change in treatment intensity. This we estimate by:

$$\hat{\theta} \ = \frac{\sum_{i=1}^{N} S_i(\Delta Y_i - g_{\hat{\lambda}}(D_{1i}, 0))}{\sum_{i=1}^{N} |\Delta D_i|} \tag{3}$$

Where S_i is the sign of the change in treatment intensity for unit i (if positive then the birdsong metric increased and vice versa), ΔY_i is the change in the outcome variable between the two periods for unit i, $|\Delta D_i|$ is the absolute change in treatment intensity, and $g_{\hat{\lambda}}(D_{1i}, 0))$ is the imputed counterfactual trend the unit would have experienced if they had not changed treatment intensity. That is, $g_{\hat{\lambda}}(D_{1i}, 0))$ is the imputed value of Y_{it} when treatment in period 1 is D_{1i} and there is no change in treatment in period 2. $g_{\hat{\lambda}}(D_1, \delta))$ means treatment in period 1 is D_1 and treatment in period 2 is $D_1 + \delta$.

For a given level of treatment the function can be separated into the non-treatment change (trend) and the change due to the change in treatment, given the parallel trend assumption (de Chaisemartin, D'Haultfœuille and Vazquez-Bare, 2024):

$$g(d_1,\delta) = \underbrace{E[Y_2(d_1) - Y_1(d_1)|\mathbf{x}, D_1 = d_1]}_{\text{Trend Effect}} \tag{4}$$
$$+ \underbrace{\delta E\left[\frac{Y_2(d_1+\delta) - Y_2(d_1)}{\delta}\middle|\mathbf{x}, D_1 = d_1, \Delta D = \delta\right]}_{\text{Change in Treatment Effect}}$$

We assume the form of the $g_{\hat{\lambda}}(d_1,\delta)$ function with a parametric linear model:

$$E[Y_2(d_1) - Y_1(d_1)|\mathbf{x}, D_1 = d_1] = \lambda_1 + \lambda_2 d_1 + \mathbf{x}_i\boldsymbol{\psi_1} + d_1 \cdot \mathbf{x}_i\boldsymbol{\psi_2}$$

$$\delta E\left[\frac{Y_2(d_1+\delta) - Y_2(d_1)}{\delta}\middle|\mathbf{x}, D_1 = d_1, \Delta D = \delta\right] = \tag{5}$$
$$\lambda_3\delta + \lambda_4 d_1 \cdot \delta + \lambda_5\delta^2 +$$
$$\delta \cdot \mathbf{x}_i\boldsymbol{\psi_3} + d_1 \cdot \delta \cdot \mathbf{x}_i\boldsymbol{\psi_4} + \delta^2 \cdot \mathbf{x}_i\boldsymbol{\psi_5}$$

We estimate these terms by a regression in the form:

$$\Delta Y_i = \lambda_1 + \lambda_2 d_1 + \mathbf{x}_i\boldsymbol{\psi_1} + d_1 \cdot \mathbf{x}_i\boldsymbol{\psi_2} + \tag{6}$$
$$\lambda_3\delta + \lambda_4 d_1 \cdot \delta + \lambda_5\delta^2 +$$
$$\delta \cdot \mathbf{x}_i\boldsymbol{\psi_3} + d_1 \cdot \delta \cdot \mathbf{x}_i\boldsymbol{\psi_4} + \delta^2 \cdot \mathbf{x}_i\boldsymbol{\psi_5}$$

The trend without a change in treatment (i.e. $g_{\hat{\lambda}}(D_1, 0)$) is represented by the first four terms in the above equation. Therefore we use those four terms as the counterfactual trend the unit would have experienced but for the change in treatment. We then plug those values into equation III.

REFERENCES

Abrahams, Carlos, Bob Ashington, Ed Baker, Tom Bradfer-Lawrence, Ella Browning, Jonathan Carruthers-Jones, Jennifer Darby, Jan Dick, Alice Eldridge, David Elliott, Becky Heath, Paul Howden-Leach, Alison Johnston, Alexander Lees, Christoph Meyer, Usue Ruiz Arana, Siobhan Smyth, and Oliver Metcalf. 2023. *Good practice guidelines for long-term ecoacoustic monitoring in the UK.*

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna. 2024. "Difference-in-differences with a Continuous Treatment."

de Chaisemartin, Clément, Xavier D'Haultfœuille, and Gonzalo Vazquez-Bare. 2024. "Difference-in-Differences Estimators with Continuous Treatments and No Stayers."

Hansen, Ben B., and Jake Bowers. 2008. "Covariate Balance in Simple, Stratified and Clustered Comparative Studies." *Statistical Science*, 23(2): 219–236. Publisher: Institute of Mathematical Statistics.

Morrison, Catriona A., Simon J. Butler, Robert A. Robinson, Jacquie A. Clark, Juan Arizaga, Ainars Aunins, Oriol Baltà, Jaroslav Cepák, Tomasz Chodkiewicz, Virginia Escandell, Ruud P. B. Foppen, Richard D. Gregory, Magne Husby, Frédéric Jiguet, John Atle Kålås, Aleksi Lehikoinen, Åke Lindström, Charlotte M. Moshøj, Károly Nagy, Arantza Leal Nebot, Markus Piha, Jiří Reif, Thomas Sattler, Jana Škorpilová, Tibor Szép, Norbert Teufelbauer, Kasper Thorup, Chris van Turnhout, Thomas Wenninger, and Jennifer A. Gill. 2021. "Covariation in population trends and demography reveals targets for conservation action." *Proceedings of the Royal Society B: Biological Sciences*, 288(1946): 20202955. Publisher: Royal Society.

Ratcliffe, Eleanor. 2021. "Sound and Soundscape in Restorative Natural Environments: A Narrative Literature Review." *Frontiers in Psychology*, 12. Publisher: Frontiers.

Ratcliffe, Eleanor, Birgitta Gatersleben, and Paul T. Sowden. 2016. "Associations with bird sounds: How do they relate to perceived restorative potential?" *Journal of Environmental Psychology*, 47: 136–144.

Topp, Christian Winther, Søren Dinesen Østergaard, Susan Søndergaard, and Per Bech. 2015. "The WHO-5 Well-Being Index: A Systematic Review of the Literature." *Psychotherapy and Psychosomatics*, 84(3): 167–176.