



Article

Power Analysis for Human Melatonin Suppression Experiments

Manuel Spitschan ^{1,2,3,*,†} , Parisa Vidafar ^{4,5} , Sean W. Cain ⁵ , Andrew J. K. Phillips ⁵ and Ben C. Lambert ^{6,*,†}

- ¹ Department of Health and Sport Sciences, TUM School of Medicine and Health, Technical University of Munich, 80992 Munich, Germany
- ² TUM Institute for Advanced Study (TUM-IAS), Technical University of Munich, 85748 Garching, Germany
- ³ Max Planck Research Group Translational Sensory and Circadian Neuroscience, Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany
- ⁴ Faculty of Medicine and Health, Central Clinical School, University of Sydney, Sydney, NSW 2006, Australia; parisa.vidafar@sydney.edu.au
- ⁵ Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Clayton, VIC 3800, Australia; sean.cain@monash.edu (S.W.C.); andrew.phillips@monash.edu (A.J.K.P.)
- ⁶ Department of Statistics, University of Oxford, Oxford OX1 3LB, UK
- * Correspondence: manuel.spitschan@tum.de (M.S.); ben.lambert@stats.ox.ac.uk (B.C.L.)
- † These authors contributed equally to this work.

Abstract: In humans, the nocturnal secretion of melatonin by the pineal gland is suppressed by ocular exposure to light. In the laboratory, melatonin suppression is a biomarker for this neuroendocrine pathway. Recent work has found that individuals differ substantially in their melatonin-suppressive response to light, with the most sensitive individuals being up to 60 times more sensitive than the least sensitive individuals. Planning experiments with melatonin suppression as an outcome needs to incorporate these individual differences, particularly in common resource-limited scenarios where running within-subjects studies at multiple light levels is costly and resource-intensive and may not be feasible with respect to participant compliance. Here, we present a novel framework for virtual laboratory melatonin suppression experiments, incorporating a Bayesian statistical model. We provide a Shiny web app for power analyses that allows users to modify various experimental parameters (sample size, individual-level heterogeneity, statistical significance threshold, light levels), and simulate a systematic shift in sensitivity (e.g., due to a pharmacological or other intervention). Our framework helps experimenters to design compelling and robust studies, offering novel insights into the underlying biological variability in melatonin suppression relevant for practical applications.

Keywords: melatonin suppression; non-visual effects of light; statistical analysis; power analysis; experimental design



Citation: Spitschan, M.; Vidafar, P.; Cain, S.W.; Phillips, A.J.K.; Lambert, B.C. Power Analysis for Human Melatonin Suppression Experiments. *Clocks&Sleep* **2024**, *6*, 114–128. <https://doi.org/10.3390/clockssleep6010009>

Academic Editor: Andrew Coogan

Received: 29 November 2023

Revised: 7 February 2024

Accepted: 10 February 2024

Published: 26 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Light exposure has a profound impact on human physiology and behaviour. In addition to enabling vision, light elicits a range of physiological and behavioural responses including the acute suppression of nocturnal melatonin production by light and shifting of the endogenous circadian rhythm [1,2]. These effects, often summarized under the umbrella term ‘non-visual’ effects of light [3], are mediated by a pathway connecting the eye to the hypothalamus. More specifically, the suprachiasmatic nuclei (SCN) receive retinofugal input, predominantly from a subset of retinal ganglion cells which are photosensitive in the absence of input from the cones and rods, the canonical photoreceptors underlying visual function. This photosensitivity is owed to the expression of the short-wavelength-sensitive photopigment melanopsin [4–7]. There is converging and convincing evidence linking the spectral sensitivity of melanopsin to the in vivo sensitivity of circadian and neuroendocrine responses to light [8–13].

Most evidence for the melatonin-suppressive effects of light is generated in laboratory experiments, in which participants are exposed to carefully controlled illumination (for

examples, see [14–17]), while their saliva, plasma, or urine is collected for later melatonin assay [18,19]. As these experiments can be resource-intensive in terms of participant burden, processing costs, and staff to run a specific multi-hour (and potentially multi-day) experimental protocols, an important consideration in designing experiments is statistical power.

Non-visual sensitivity to light exhibits large individual differences [20,21], as demonstrated by a recent study by Phillips et al. [22], in which participants were exposed to overhead fluorescent illumination at different illuminance levels on different evenings (10 lx, 100 lx, 200 lx, 400 lx, 1000 lx, 2000 lx). During this time their melatonin secretion was measured, showed substantial individual differences, with the most sensitive individual being up to 60 times more sensitive than the least sensitive individual in their sample.

These sizeable individual differences require special attention in designing experiments to obtain compelling evidence, particularly when it is not feasible to run extensive within-subjects experiments sampling the same individuals multiple times under multiple illuminances (thereby minimising individual differences). Here, we present a novel framework for facilitating power analyses for human melatonin suppression experiments to optimize the choice of experimental illumination levels.

2. Results

2.1. Overview

We developed a Bayesian statistical model of a virtual laboratory melatonin suppression experiment [22]. The model can be thought of as representing the melatonin suppression versus photopic illuminance [lx] curves (henceforth termed “dose-response curves”) across the population from which the original study participants were sampled. The model is stochastic, meaning that each virtual individual drawn from it will likely have a different dose-response curve. Figure 1 shows four replicates of a virtual experiment comprising $n = 41$ participants in each case (the same sample size as the estimates in [22]). There are two types of stochasticity inherent in the model: individual-level variation in dose-response curve parameters; and measurement error when recording an individual’s melatonin suppression level at a single discrete illuminance value. These two sources of uncertainty are evident in Figure 1 by the variation in dose-response curves between subjects and the variation in measurements (black points) around them.

The assumptions underpinning our model of virtual experiments are detailed in Section 2.2. The ability to generate such virtual experiments allows us to conduct a series of *in silico* experiments to estimate the statistical power for a series of different laboratory study designs. The various study types are outlined in Section 2.3. As part of this work, we have created an open-source R package called *melluxdrc* [23] that wraps the functionality required to generate virtual experiments and to perform statistical power calculations. To facilitate power calculations for those designing experiments, we have also conducted a series of calculations and made those results freely available via an online R Shiny application called the *mellux-app* [24], which we describe briefly in Section 5.3.

2.2. A Model of Virtual Experiments

Our model of virtual experiments comprises two elements: (a) a population model of dose-response curves and (b) a measurement error model that statistically represents the various noise factors that influence measurements of melatonin suppression at a particular lux level.

The population model of dose-response curves is based on parameter estimates presented for $n = 41$ participants in [22]. In [22], dose-response data for $n = 55$ participants were obtained in a within-subjects protocol, where participants were exposed to a dim control (<1 lux) and five other experimental light levels (10, 30, 50, 100, 200, 400, 2000 lux) for 5 h in the evening. Using these data, melatonin suppression values, $s(x, i)$, were obtained at

each experimental light level, x for each participant i . The series of melatonin suppression values for each participant were then modelled using a logistic-type curve of the form:

$$s(x, i) = 1 - \frac{1}{1 + \left(\frac{\log_{10} x}{a_i} \right)^{b_i}}, \quad (1)$$

which has the property that melatonin suppression increases towards 100% as illuminance increases without bound; here, $b_i > 0$ controls the shape of each individual's dose-response curve, and $a_i = \log_{10} ED_{50}$ for that individual. Here, ED_{50} is the dose corresponding to a melatonin suppression value of 50%.

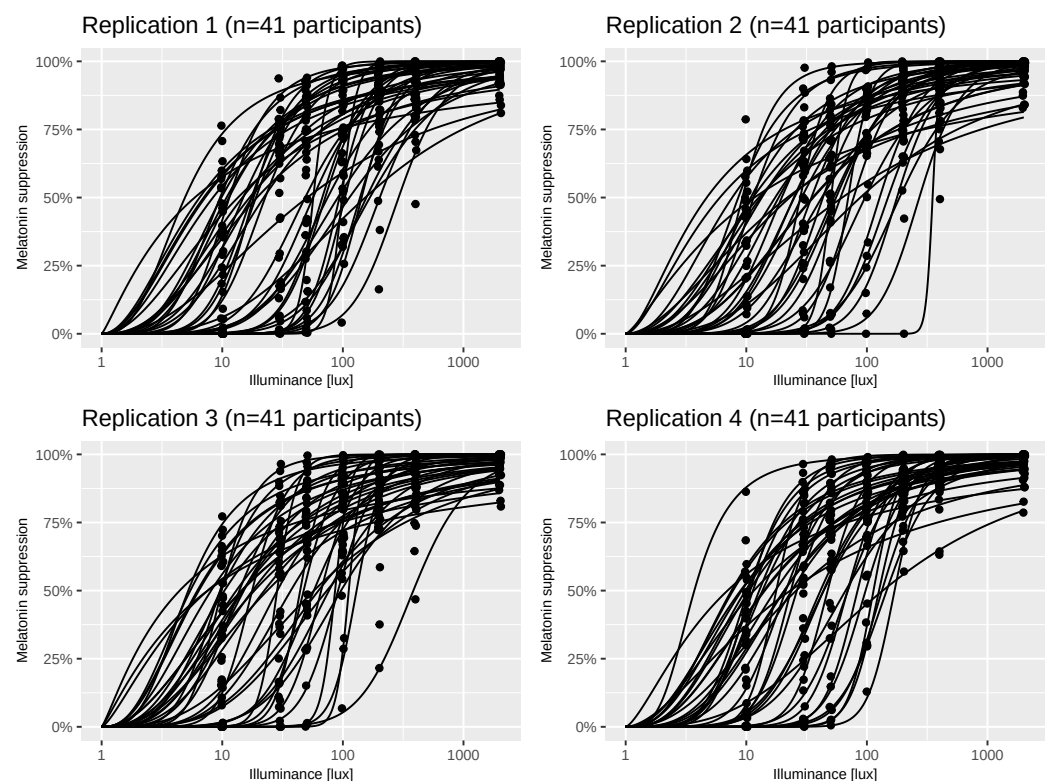


Figure 1. Example virtual experiments. Each panel shows individual dose-response curves and measurements (black points) of melatonin suppression at a series of illuminance values for $n = 41$ participants generated using the `virtual_experiment` ($n = 41$) function from the `melluxdrc` R package. The assumptions underpinning each of these panels are described in Section 2.2.

The values used for our analysis were the (a_i, b_i) estimates for the $n = 41$ participants previously reported in which a reliable estimate of the dose-response curve was obtained (less than 1 log-unit 95% confidence interval for the estimate of a_i): each set characterising the dose-response curve as in Equation (1). From here onwards, we refer to this set of estimates as the raw dose-response estimates. Using these values, we sought to create a statistical population model that represented these collection of dose-response curves which, crucially, could be sampled from to obtain a dose-response curve for a virtual individual.

To do so, we built a statistical distribution representing the raw estimates. To represent this bivariate distribution, we used kernel density estimation using the `bkde` function from the `KernSmooth` R package [25] to approximate the empirical distribution over a_i values. We then modelled the conditional distribution of $\log b_i$ conditional on a_i through a regression equation:

$$\log b_i \sim \text{normal}(\alpha + \beta a_i, \sigma_0 + \sigma_1 a_i), \quad (2)$$

where the regression equation allows heteroscedasticity, represented by a standard deviation that increases linearly with a_i (since $\sigma_0 > 0$ and $\sigma_1 > 0$). This model was estimated using a Bayesian framework, meaning that priors were set on the parameters. The priors chosen were uninformative, allowing a wide range of possible relationships and are shown in Table 1. The model was fitted using Markov chain Monte Carlo (MCMC) through Stan's NUTS algorithm [26], using 4 Markov chains with 4000 iterations per chain, with 2000 of each chain's iterations discarded as warm-up; finally, the post-warm-up iterations were thinned by a factor of 2. MCMC convergence was diagnosed through $\hat{R} < 1.01$ and having bulk- and tail-ESS values above 400 [27]. Posterior predictive checks of model fit (see, for example, [28]) indicated that the model was a reasonable fit to the data (Figure 2). The Stan file and all materials needed to reproduce this analysis are available at [29].

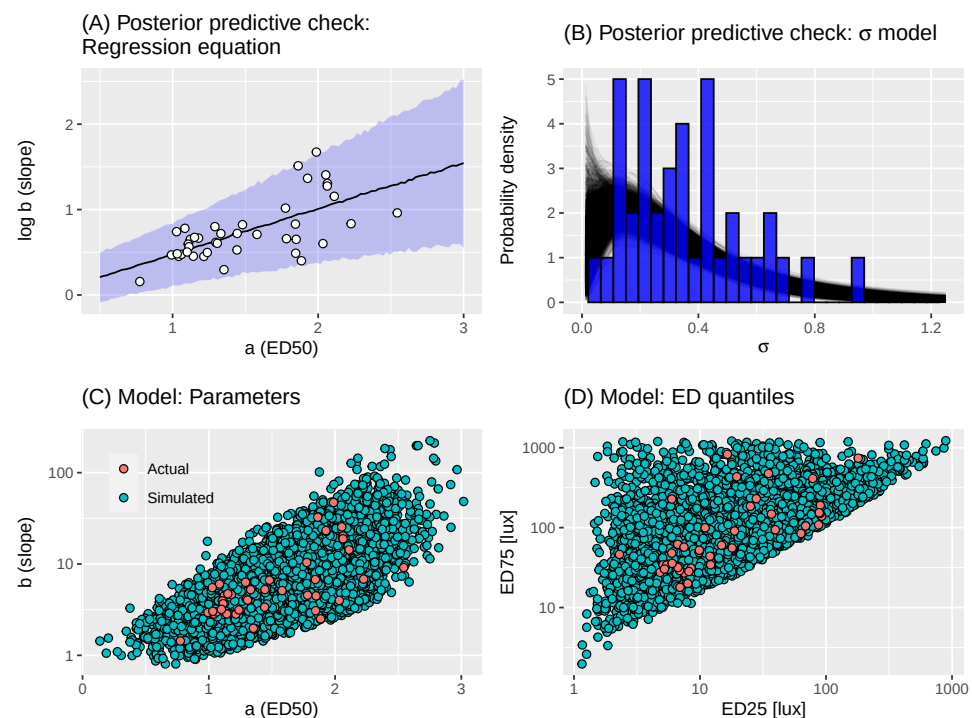


Figure 2. (A) Posterior predictive check: regression equation. Plot shows a graphical check of the model fit of Equation (2). The individual points show the estimates of (a_i, b_i) parameters provided by the authors of [22]. The uncertainty ribbon indicates the 2.5%-97.5% posterior predictive quantiles; the black line indicates the 50% posterior predictive quantile. (B) Posterior predictive check: σ_i model. Plot shows a graphical check of the model fit of Equation (4). Each black line represents a gamma density function corresponding to particular posterior samples of the parameters. The blue bars indicate the values of σ_i estimated by the root-finding algorithm. (C) Assessing virtual individual generation: dose-response parameters. Each orange point represents a draw of (a, b) parameters (in Equation (1)) obtained via Algorithm 1: here, we show 25,000 such estimates; each green point represents raw estimates from [22]. (D) Assessing virtual individual generation: ED_x quantiles. Each orange point represents the (ED_{25}, ED_{75}) values correspond to a draw of (a, b) parameters (in Equation (1)) obtained via Algorithm 1: here, we show 25,000 such estimates; each green point represents the ED_x quantiles corresponding to the raw estimates from [22].

Table 1. Priors. Shows the priors used on the parameters of the Bayesian models that were estimated.

Model	Prior
Equation (1)	$\alpha \sim \text{normal}(0, 1)$
Equation (1)	$\beta \sim \text{normal}(0, 1)$
Equation (1)	$\sigma_0 \sim \text{Cauchy}^+(0, 1)$
Equation (1)	$\sigma_1 \sim \text{Cauchy}^+(0, 1)$
Equation (4)	$a \sim \text{Cauchy}(0, 1)$
Equation (4)	$b \sim \text{Cauchy}(0, 1)$

To sample parameters characterising a dose-response curve for a virtual individual, we then used the approach described in Algorithm 1. Briefly, the first section of the algorithm draws a from the kernel density estimate of the empirical distribution. Then, Equation (2) is used to draw a value of $\log b$ conditional on a . The final section of the algorithm calculates the ED_{25} and ED_{75} corresponding to the sampled a and b values. If either of these are more extreme than thresholds derived from the corresponding ED values from the estimates from [22], those (a, b) values are rejected and the function is called again. This last step ensures that the dose-response curves obtained for virtual individuals are not far more extreme than those witnessed in the raw estimates. This approach is able to generate samples of (a, b) parameters that encompass the distribution of raw estimates (Figure 3). Accordingly, the corresponding ED_{25} and ED_{75} values generated by this process were also a reasonable fit to those corresponding to the raw estimates (Figure 3).

Algorithm 1 Virtual individual generation. Takes as input posterior draws of $\alpha, \beta, \sigma_0, \sigma_1$ in Equation (2)

```

1: procedure VIRTUALINDIVIDUAL( $\alpha, \beta, \sigma_0, \sigma_1$ )
2:   Sample  $a$  from a kernel density approximation of empirical distribution using
   inverse transform sampling
3:   Uniformly sample a set of random  $(\alpha, \beta, \sigma_0, \sigma_1)$  values from posterior draw set
4:   Use Equation (2) to draw a value of  $\log b$  (and hence  $b$ )
5:   Use  $ed(x) = 10^{a(-1+1/(1-x))^{1/b}}$  to calculate  $ED_{25}$  ( $x = 0.25$ ) and  $ED_{75}$  ( $x = 0.75$ ) for
   that individual
6:   If  $ED_{25}$  is less than half the minimum  $ED_{25}$  in estimates from [22], reject  $(a, b)$  and
   call VirtualIndividual
7:   If  $ED_{75}$  is greater than 1.5 times the maximum  $ED_{75}$  in estimates from [22], reject
    $(a, b)$  and call VirtualIndividual
8:   return  $(a, b)$  which characterise a dose-response curve
9: end procedure

```

As part of our virtual population, we also developed an algorithm to allow users to simulate a reduction in individual heterogeneity in the population. To do so, we introduced a parameter, $0 \leq \eta \leq 1$ that modulates the level of individual heterogeneity in the population: here, $\eta = 0$ indicates that virtual individuals sampled from the population all have the same dose-response curve near the population median; $\eta = 1$ indicates that the virtual individuals are drawn from the unrestricted population model (i.e., by the same process described in Algorithm 1). The process used to sample virtual individuals from the population is provided in Algorithm 2. Figure 4 shows how reducing the individual-level heterogeneity using this approach results in a tighter spread of dose-response curves.

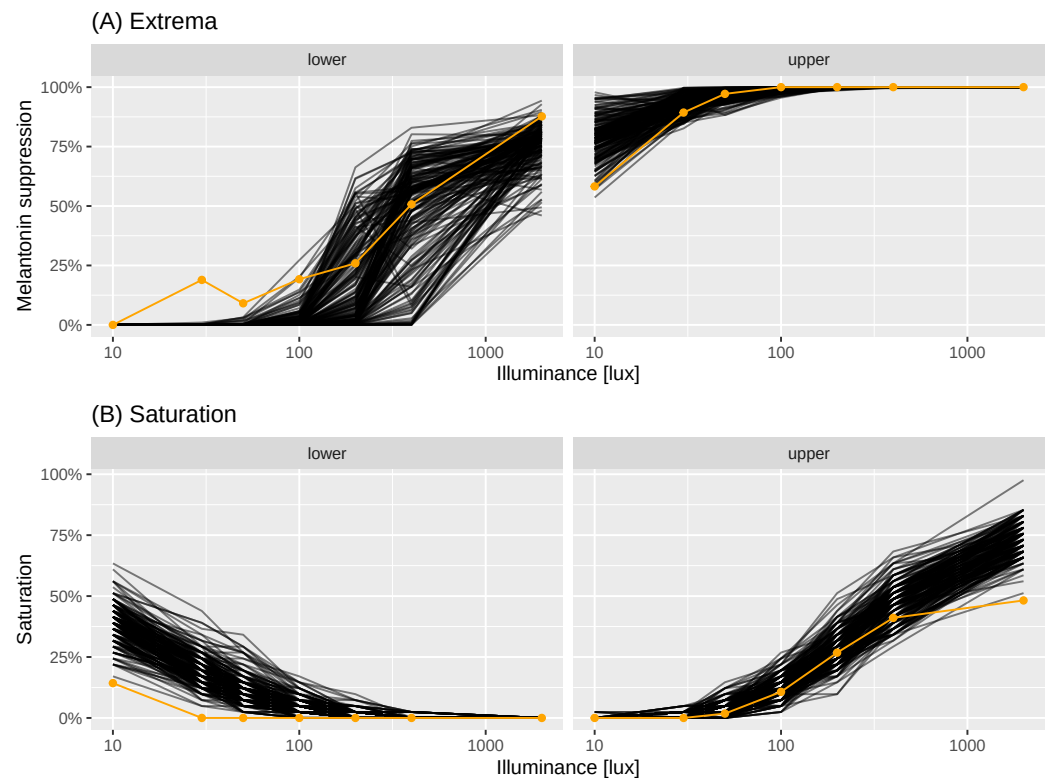


Figure 3. Virtual experiment check: saturations. Panels correspond to lower (percentage of observations < 5%) and upper (percentage of observations > 95%) saturations. Each black line corresponds to saturations generated from a single virtual experiment of sample size $n = 41$. Each orange point corresponds to the real saturation.

Algorithm 2 Virtual individual generation: reduced individual variance. Takes as input posterior draws of $\alpha, \beta, \sigma_0, \sigma_1$ in Equation (2) and η

- 1: **procedure** VIRTUALINDIVIDUALREDUCEDVARIANCE($\eta, \alpha, \beta, \sigma_0, \sigma_1$)
- 2: For all α_i posterior draws, shrink towards the grand mean: $\alpha_i = \alpha_i + (\bar{\alpha} - \alpha_i)(1 - \eta)$ where $\bar{\alpha}$ is the posterior mean
- 3: For all β_i posterior draws, shrink towards the grand mean: $\beta_i = \beta_i + (\bar{\beta} - \beta_i)(1 - \eta)$ where $\bar{\beta}$ is the posterior mean
- 4: Uniformly sample a set of random (α, β) values from the variance-reduced posterior draw set
- 5: Uniformly sample a set of random (σ_0, σ_1) values from the posterior draw set
- 6: Sample a' from a kernel density approximation of empirical distribution using inverse transform sampling (see [30])
- 7: Calculate $a = a' + (a_{50} - a')(1 - \eta)$ where a_{50} indicates the median value of the empirical distribution
- 8: Calculate $\sigma = \eta(\sigma_0 + \sigma_1 a)$
- 9: Draw a value of $\log b \sim \text{normal}(\alpha + \beta a_i, \sigma)$
- 10: Use $ed(x) = 10^{a(-1+1/(1-x))^{1/b}}$ to calculate ED_{25} ($x = 0.25$) and ED_{75} ($x = 0.75$) for that individual
- 11: If ED_{25} is less than half the minimum ED_{25} in estimates from [22], reject (a, b) and call VirtualIndividualReducedVariance
- 12: If ED_{75} is greater than 1.5 times the maximum ED_{75} in estimates from [22], reject (a, b) and call VirtualIndividualReducedVariance
- 13: **return** (a, b) which characterize a dose-response curve
- 14: **end procedure**

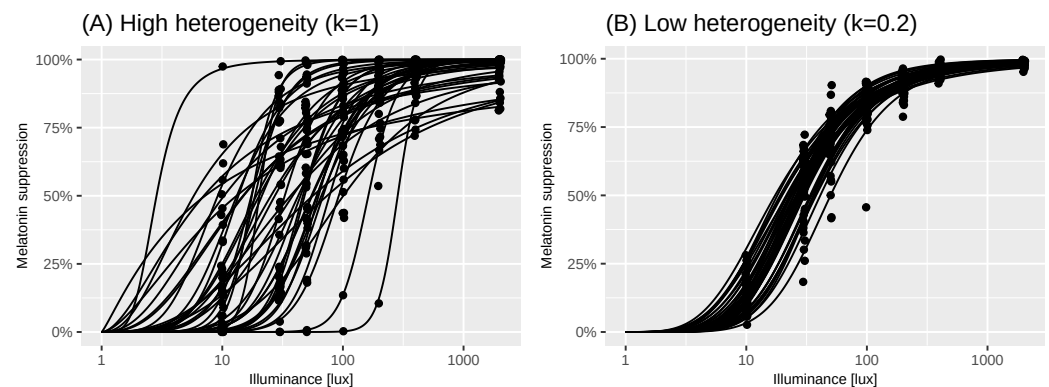


Figure 4. Effect of reduction in individual variance. Panel A shows individual dose-response curves and measurements (black points) of melatonin suppression at a series of illuminance values for $n = 41$ participants generated using `virtual_experiment` ($n = 41$) function from the `melluxdrc` R package. Panel B shows the same but assuming a reduction in individual variance given by $\eta = 0.2$ as given by `virtual_experiment` ($n = 41$, `individual_variation_level` = 0.2). The assumptions underpinning each of these panels are described in Section 5.

As is seen in Figure 2 in [22], the measured melatonin suppression values at discrete illuminance values exhibited often considerable variation around the best fit lines. Using the Root-Mean-Square-Errors (RMSEs) for each fit in for each individual, we developed an error model representing measurement variability for each individual. To do so, we assumed that measurement noise was additive on the logit scale to ensure that measurements could not fall outside of the $[0, 1]$ range. That is,

$$\text{logit } \tilde{s}(x, i) \sim \text{normal}(\text{logit } s(x, i), \sigma_i), \quad (3)$$

where $\text{logit}(z) := \log z / (1 - z)$, and $\tilde{s}(x, i)$ represents the measured melatonin suppression at illuminance x for individual i . Here, $\sigma_i > 0$ is an individual-specific value quantifying measurement noise.

We aimed to determine the value of σ_i that generated measurement noise resulting in a corresponding RMSE value close to that observed for individual i . To do so, we used an approximate estimation approach where, in each iteration, we simulated data using Equations (1) and (3), and compared the simulated RMSE value to the truth. Specifically, we used the known (estimated) (a_i, b_i) values for individual i to simulate a dose-response curve given by Equation (1); we then generated measurements at the discrete illuminance values used in [22] using Equation (3) with a particular σ_i . For each such “experiment”, we calculated an RMSE value. For a given σ_i value, we repeated the experiment 100 times, and calculated an average RMSE value across all replicates. The difference between this average simulated value and the true RMSE value was then used as the target for the one-dimensional root-finding algorithm `uniroot` available in R: the output of this algorithm was, hence, a value of σ_i for each individual.

In order to generate measurements for virtual individuals, we needed an approach to generate potential σ_i values for unseen individuals. To do so, we assumed that the individual σ_i values were drawn from the following population process:

$$\sigma_i \sim \text{gamma}(c, d), \quad (4)$$

where $c > 0$ and $d > 0$. We estimated the parameters in Equation (4) using the values of σ_i estimated by the root-finding algorithm. The model was estimated in a Bayesian paradigm, requiring that we set priors on the parameters. Here, we chose uninformative priors on the parameters, which are specified in Table 1. The model was fitted using Markov chain Monte Carlo (MCMC) through Stan’s NUTS algorithm [26] using 4 Markov chains with 2000 iterations per chain, with 1000 of each chain’s iterations discarded as warm-up.

The same criteria were used to diagnose model convergence as for Equation (2). Posterior predictive checks of model fit indicated that the model was a reasonable fit to the data (Figure 2).

With both elements of the virtual experiment model in place – the population model of dose-response experiments and the measurement error model – we sought to assess how well the resultant model represented the data collected in [22]. This was tested using the minimum and maximum melatonin suppression at each measured illuminance value (the “extrema”); and the percentage of observations which were extreme at each lux: either below 5% lux or above 95% melatonin suppression (the lower and upper “saturation values”). We generated 200 replicate virtual experiments, which each had the same number of individuals ($n = 41$) as the estimates we were provided with. For each experiment and illuminance value, we calculated simulated extrema and saturation values to compare with the real measurements. In Figure 3, we compare the modelled extrema (black lines) with the real values (orange points and line): in the left-hand plot, we compare the minima; in the right plot, we compare the maxima. This indicates that, for moderate-high illuminance values, the modelled and real values were in good accordance. At the lower illuminance values, the modelled values tended to be more extreme (either closer to 0% for the minima; or closer to 100% for the maxima) than the real data. In Figure 3, we performed the same comparison but for the levels of simulated and actual saturations. In this case, the upper saturation values between the modelled and real data were reasonable, apart from at the highest illuminance value measured. The lower saturations exhibited the same pattern as for the extrema, in that the simulated data were more extreme than the actual.

The discrepancies between the modelled and actual data could be due to (a) assumptions around our model for the population of dose-response curves, (b) assumptions around our measurement process model and/or (c) issues with the original logistic models used to model dose-response curves in [22]. Our checks for our model of dose-response curves (Figure 3) indicate no issue with representing the original estimates. It is possible that the logit-normal noise process we assume in Equation (3) results in greater variation than seen in reality. Visual inspection of Figure 2 in Phillips et al. [22], however, shows that, for a number of individuals, the logistic dose-response curve is downward-biased for low illuminance levels, potentially indicating that a two-parameter logistic is inappropriate in this extreme. Without the raw experimental data, however, it is not possible to determine the exact cause of the discrepancy between modelled and real life data. Despite these differences, the overall correspondence between the model and data is good, and, as such, it provides a reasonable basis to determine statistical power for a variety of different experimental settings.

2.3. Power Calculations

Having developed a virtual experiment model (see Section 2.2), we sought to use it to help inform experimental design. In particular, we considered two classes of experiment: in illuminance comparison experiments, the aim is to quantify whether population melatonin suppression differs across two measured illuminance values; in intervention experiments, the aim is to estimate the effect size for an intervention that systematically changes the dose-response curves (i.e., changes light sensitivity). Within each experimental class, we considered two possible experimental designs: within-subject designs, where the same participants are measured twice; or between-subject designs, where two separate groups of participants are measured once each. In all cases, we used our model of virtual experiments to generate data that mimics the types of laboratory experiments described here, which was then used to determine statistical power (Figures 5 and 6).

In two-level experiments, individual melatonin suppressions are measured at two photopic illuminance levels: x_1 and x_2 . In within-subject experiments, each individual is measured twice resulting in N paired observations: $\mathcal{D} = \{\tilde{s}(x_1, i), \tilde{s}(x_2, i)\}^{1:N}$. In between-subject experiments, there are $2N$ individuals, with N measured at x_1 and another N at x_2 , resulting in a dataset: $\mathcal{D} = \{\{\tilde{s}(x_1, i)\}^{1:N} \& \{\tilde{s}(x_2, i)\}^{N+1:2N}\}$. In these experiments,

the aim is to determine that there is a significant difference in the suppression response between measurements taken at x_1 and x_2 . To do so, we use a t-test which is either paired (for within-subject designs) or independent (for between-subject designs). We considered all possible pairs of different illuminance values from the set $\{10, 30, 50, \dots, 1970, 1990\}$. For each pair, we also considered a range of sample sizes from $N = 10, 20, \dots, 90, 100$ and a range of possible individual heterogeneity values: $\eta = 0, 0.2, 0.4, 0.6, 0.8, 1.0$. For each combination of these parameters, we calculated the power to detect significant differences of the correct sign using test sizes of 1%, 5% and 10%.

We show visualisations of the outcomes for these analyses in Figures 6 and 7. Figure 6 shows the statistical power as a function of participant level heterogeneity for between- and within-subjects designs. The same principle of power but for different illuminances levels is shown in Figure 7. The results are clear: within-subjects designs beat between-subjects designs.

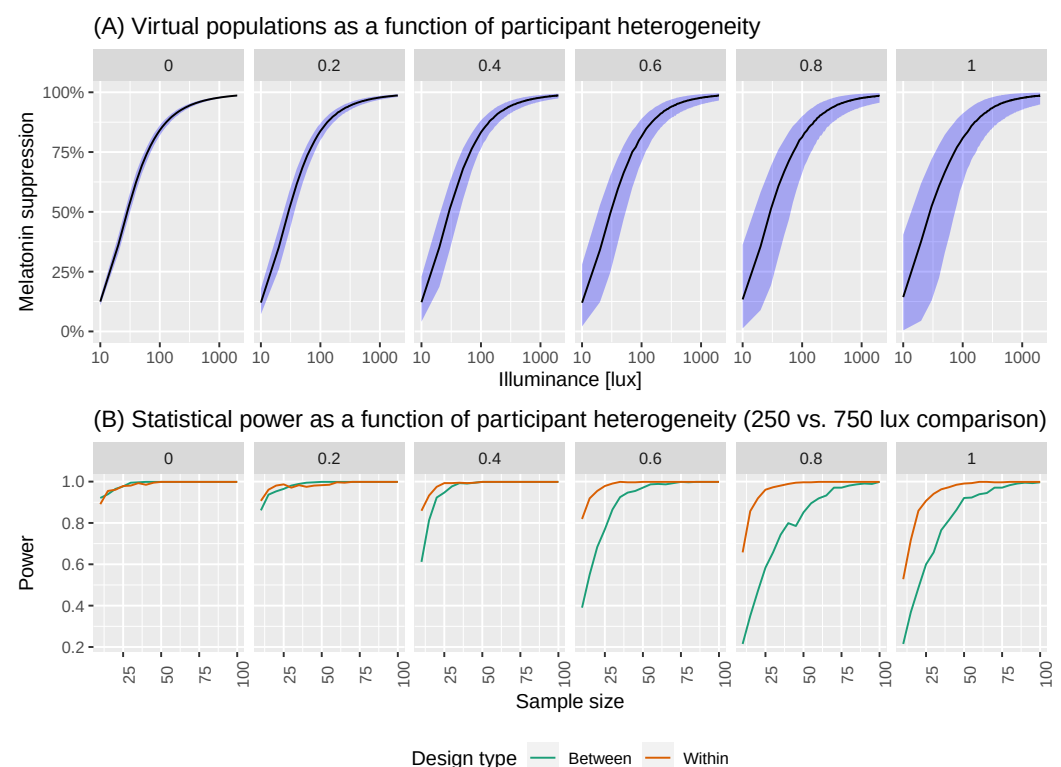


Figure 5. (A) Virtual populations as a function of participant heterogeneity. Here, we are simulating the group dose-response curves assuming different levels of participant heterogeneity. (B) Statistical power as a function of participant heterogeneity.

In intervention experiments, we assume interventions that shift the ED_{50} away from the baseline level. In within-subject study designs, individuals are assumed to have their melatonin suppressions measured at a particular illuminance level (in lux) x once before the intervention occurs and once afterwards, resulting in a dataset: $\mathcal{D} = \{\tilde{s}(x, i), \tilde{s}'(x, i)\}^{1:N}$ where unprimed variables indicate natural levels and primed indicate intervention levels. For the between-subject design, two different sets of individuals were measured: a “baseline” group and an “intervention” group whose ED_{50} values are shifted. This resulted in a dataset $\mathcal{D} = \{\{\tilde{s}(x, i)\}^{1:N} \& \{\tilde{s}'(x, i)\}^{N+1:2N}\}$. As for the illuminance comparison experiments, we used t-tests to compare the natural and intervention melatonin suppressions: again using paired-tests for the within-subject design and independent tests for the between-subject design. We considered all illuminance values from the set $x = \{10, 30, 50, \dots, 1970, 1990\}$ across the same set of sample sizes and individual heterogeneities as for the comparison experiments. Additionally, we considered natural ED_{50}

multipliers taking values $\chi = \{0.2, 0.4, \dots, 1.8, 2.0\}$, where, for example, $\chi = 0.2$ means the natural ED_{50} of individuals are reduced by a factor of 5. For each combination of these parameters, we calculated the power to detect significant differences of the correct sign using test sizes of 1%, 5% and 10%.

We summarize the outcomes of these analyses in Figure 8. We find clear differences in power assuming a specific intervention effect. These analyses in Figure 8 may serve to inform sample size calculations for planning interventions affecting the impact of light on melatonin suppression.

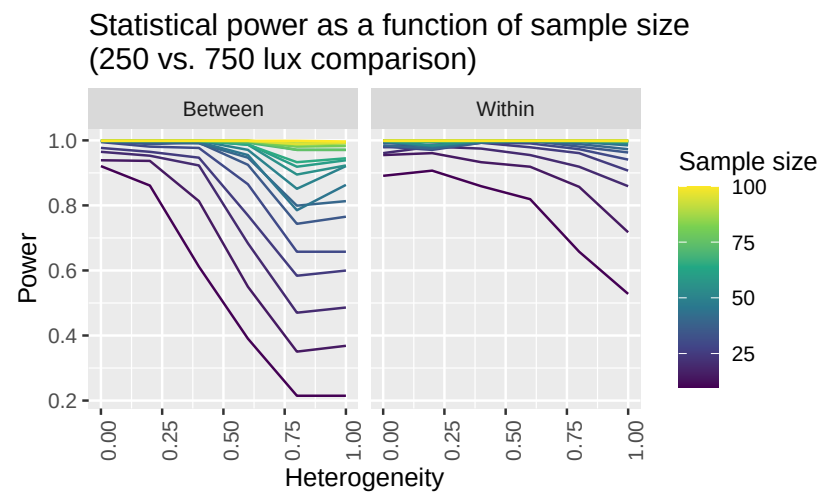


Figure 6. Statistical power for detecting a difference in melatonin suppression assuming different samples sizes and between-subjects (left panel) and within-subjects (right panel) designs.

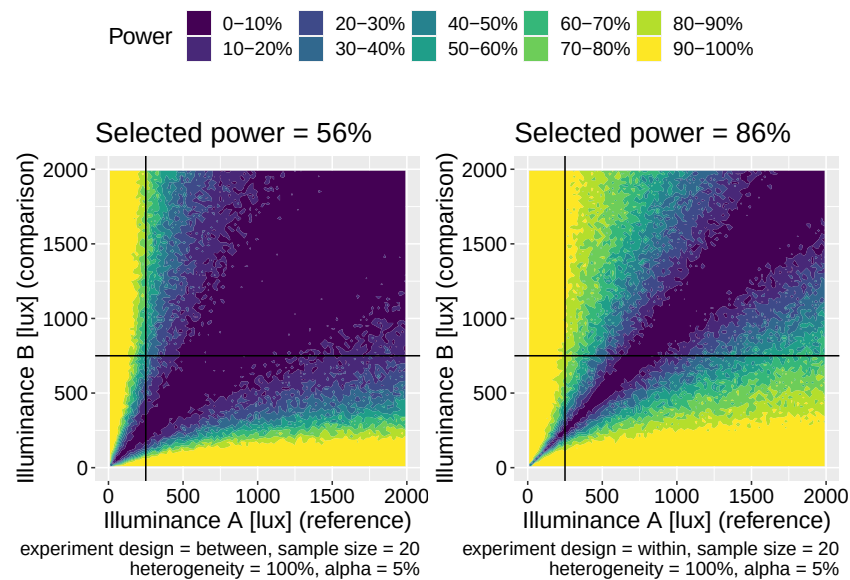


Figure 7. Power maps for two-illuminance comparison experiments for between-subjects (left panel) and within-subjects experiments (right panel). The choice of comparison light levels here is 250 and 750 lux.

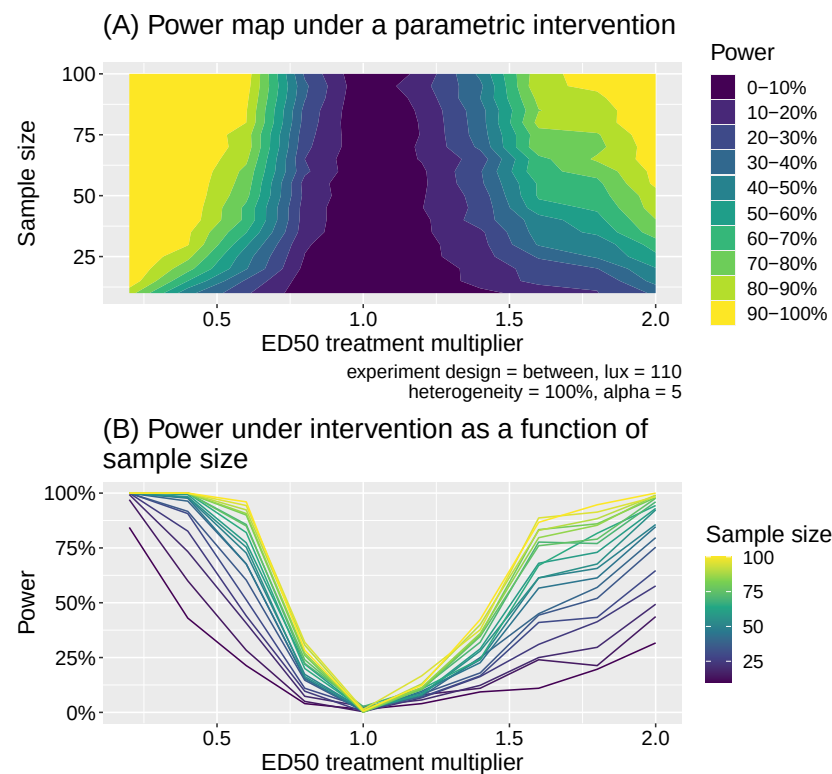


Figure 8. Panel (A) shows the power as a function of the intervention effect (expressed as a multiplier of the ED50 value) and sample size. Panel (B) shows a slice of the same data as a line plot for different discrete samples sizes.

2.4. Extension to Melanopic EDI

In the present model, we used photopic illuminance in lux as the independent variable. It is well known that melatonin suppression follows the spectral sensitivity of melanopsin [7,12]. In the experiments underlying our model [22], the spectrum did not change substantially with different intensities. As a consequence, the results here hold independent of the light quantification metrics, and using a simple scaling factor, can be expressed as melanopic equivalent daylight illuminance (mEDI) [31]. Following calculations with lux [32], the melanopic EDIs were 1.58 lx (nominal10 photopic lx), 48.24 lx (nominal100 photopic lx), 99.76 lx (nominal200 photopic lx), 206.33 lx (nominal400 photopic lx) and 1106.21 lx (nominal200 photopic lx) in the original data set.

3. Discussion

3.1. Limitations

The presented model has notable limitations stemming from the reliance on a single data set, posing challenges regarding the generalizability of our findings. The modelling approach is flexible to allow for the future integration of data sets, should these become available (see Section 2.3). The dataset was collected in a relatively homogeneous group of participants. Indeed, given the large variability in even this rather homogeneous sample, it is unlikely that more heterogeneous samples would lead to less variable results. Future work include data from diverse populations.

Due to data-sharing restrictions, only individual-level logistic fits to the data were available. This limitation raises concerns about the assumed form of the dose-response curve and the appropriateness of the chosen noise model, both of which could be better assessed with access to raw individual-level data. Additionally, the assumed form of the noise model, specifically its additive nature on the logit scale, might not be the optimal fit.

The model exhibited strengths in providing a good fit to the available data, showcasing its reliability for the bulk of the available information. Furthermore, the approach emphasizes transparency and openness by making the model accessible through reproducible code repositories and a graphical user interface. The approach's innovative aspect lies in being one of the initial attempts to construct a generative model of experimental data, allowing users to specify assumptions about individual-level variation. The use of a Bayesian approach is statistically principled by correctly incorporating various sources of uncertainty into the model's characteristics, enhancing its robustness.

3.2. Future Directions

The analysis and approach presented here provides a novel approach for power analysis in human melatonin suppression experiments. It can help to inform future studies of melatonin suppression under various manipulations, assuming individual variability, and thereby support experimental design and a principled decision about samples sizes. Importantly, the framework presented here can be updated when more datasets, and more diverse data sets, collected under similar experimental paradigms become available. The approach proposed here for modelling dose-response curves can be extended to other dose-response relationships as well. An additional consideration is the inclusion of melanopic metrology into the modelling framework, along with more sophisticated photoreceptor integration models.

4. Conclusions

Here, we developed a model of virtual melatonin suppression experiments under different light levels informed by, and fitting to, actual experimental data. We expanded our model virtual experiments by allowing the modulation of individual-level variability and implementing a virtual intervention leading to a left- or rightward shift of the dose-response curve. We provide a web applet for exploring the virtual model and using it in practical applications.

5. Methods

5.1. Sample Characteristics

The study on which we base the statistical modeling is based on a sample of young healthy adults, aged 18–30 years (mean $\pm$ SD; age of 20.8 ± 2.6 y). Data from 41 participants were included in our analyses. The empirical sample had the following characteristics [22]:

“A total of 61 participants were enrolled, of whom 3 were excluded based on actigraphy, and 2 did not complete the study beyond the baseline DLMO. Overall, 56 healthy young Caucasian adults (29 women, 27 men; 20.8 ± 2.6 y of age) completed the study. Participants were free from any medical or psychological conditions, had a BMI of 18–30 kg/m², and were not taking any medications at the time of the study. Participants had not recently traveled across time zones (1 mo per time zone, up to 3 mo) or engaged in shiftwork in the previous 12 mo. Women were naturally cycling (i.e., free from hormonal contraception) and had a regular menstrual cycle of 21–36 d in duration. Participants were healthy sleepers, reporting no subjective problems or previous diagnoses, and having a regular bedtime before 1 AM. A score of 10 or greater on the Epworth Sleepiness Scale was exclusionary, and participants were predominantly intermediate chronotypes (MEQ score of 52.7 ± 9.2). Participants had an average bedtime and waketime of 23:04 (SD = 44 min) and 07:04 (SD = 44 min), respectively. DLMO occurred on average at 21:05 (SD = 70 min), 2.22 h before bedtime. A total of nine participants wore prescription glasses at each test session.”

5.2. Reproducibility of Results

The version of `melluxdrc` used to generate the results in this paper is v1.0.0.

5.3. Shiny App

We provide an app developed in the Shiny framework which allows users to generate power analyses. The app is deployed at <https://tscnlab.shinyapps.io/mellux-app/> (accessed on 20 February 2024). Screenshots are shown in Figure 9.

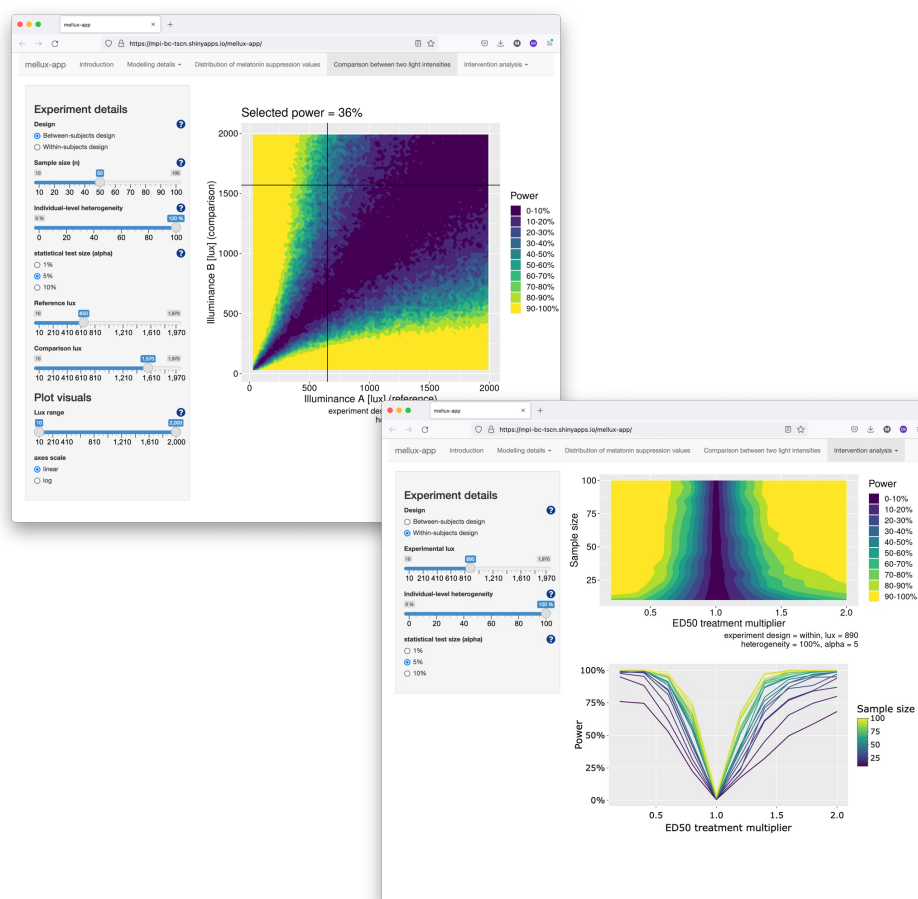


Figure 9. Screenshots of the Shiny app for exploring sample size calculations interactively.

5.4. Code and Data Availability

All code and data underlying this work is available under the MIT License. We make available the R package *melluxdrc* for generating virtual experiments [23], code to run simulations on a cluster [33], and code to fit dose response curves [29]. The code underlying the shiny app, deployed at <https://tscnlab.shinyapps.io/mellux-app/> (accessed on 20 February 2024) is available as well [24].

Author Contributions: Conceptualization, M.S. and B.C.L.; methodology, M.S. and B.C.L.; software, M.S. and B.C.L.; validation, B.C.L.; formal analysis, M.S. and B.C.L.; investigation, M.S. and B.C.L.; resources, M.S., P.V., S.W.C., A.J.K.P. and B.C.L.; data curation, P.V., S.W.C. and A.J.K.P.; writing—original draft preparation, M.S. and B.C.L.; writing—review and editing, M.S. and B.C.L.; visualization, M.S. and B.C.L.; supervision, M.S.; project administration, M.S.; funding acquisition, M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Wellcome Trust (204686/Z/16/Z, 204686/Z/16/C).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: See Section 5.4.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Blume, C.; Garbaza, C.; Spitschan, M. Effects of light on human circadian rhythms, sleep and mood. *Somnologie* **2019**, *23*, 147–156. <https://doi.org/10.1007/s11818-019-00215-x>.
2. Vetter, C.; Pattison, P.M.; Houser, K.; Herf, M.; Phillips, A.J.K.; Wright, K.P.; Skene, D.J.; Brainard, G.C.; Boivin, D.B.; Glickman, G. A review of human physiological responses to light: Implications for the development of integrative lighting solutions. *Leukos* **2021**, *18*, 387–414. <https://doi.org/10.1080/15502724.2021.1872383>.
3. Schlangen, L.J.M.; Price, L.L.A. The lighting environment, its metrology, and non-visual responses. *Front. Neurol.* **2021**, *12*, 624861. <https://doi.org/10.3389/fneur.2021.624861>.
4. Provencio, I.; Jiang, G.; De Grip, W.J.; Hayes, W.P.; Rollag, M.D. Melanopsin: An opsin in melanophores, brain, and eye. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 340–345. <https://doi.org/10.1073/pnas.95.1.340>.
5. Provencio, I.; Rodriguez, I.R.; Jiang, G.; Hayes, W.P.; Moreira, E.F.; Rollag, M.D. A novel human opsin in the inner retina. *J. Neurosci.* **2000**, *20*, 600–605.
6. Do, M.T.H. Melanopsin and the intrinsically photosensitive retinal ganglion cells: Biophysics to behavior. *Neuron* **2019**, *104*, 205–226. <https://doi.org/10.1016/j.neuron.2019.07.016>.
7. Spitschan, M. Melanopsin contributions to non-visual and visual function. *Curr. Opin. Behav. Sci.* **2019**, *30*, 67–72. <https://doi.org/10.1016/j.cobeha.2019.06.004>.
8. Brainard, G.C.; Hanifin, J.P.; Greeson, J.M.; Byrne, B.; Glickman, G.; Gerner, E.; Rollag, M.D. Action spectrum for melatonin regulation in humans: Evidence for a novel circadian photoreceptor. *J. Neurosci.* **2001**, *21*, 6405–6412.
9. Thapan, K.; Arendt, J.; Skene, D.J. An action spectrum for melatonin suppression: Evidence for a novel non-rod, non-cone photoreceptor system in humans. *J. Physiol.* **2001**, *535*, 261–267. <https://doi.org/10.1111/j.1469-7793.2001.t01-1-00261.x>.
10. Nowozin, C.; Wahnschaffe, A.; Rodenbeck, A.; de Zeeuw, J.; Hadel, S.; Kozakov, R.; Schopp, H.; Munch, M.; Kunz, D. Applying melanopic lux to measure biological light effects on melatonin suppression and subjective sleepiness. *Curr. Alzheimer Res.* **2017**, *14*, 1042–1052. <https://doi.org/10.2174/1567205014666170523094526>.
11. Prayag, A.S.; Najjar, R.P.; Gronfier, C. Melatonin suppression is exquisitely sensitive to light and primarily driven by melanopsin in humans. *J. Pineal Res.* **2019**, *66*, e12562. <https://doi.org/10.1111/jpi.12562>.
12. Brown, T.M. Melanopic illuminance defines the magnitude of human circadian light responses under a wide range of conditions. *J. Pineal Res.* **2020**, *69*, e12655. <https://doi.org/10.1111/jpi.12655>.
13. Gimenez, M.C.; Stefani, O.; Cajochen, C.; Lang, D.; Deuring, G.; Schlangen, L.J.M. Predicting melatonin suppression by light in humans: Unifying photoreceptor-based equivalent daylight illuminances, spectral composition, timing and duration of light exposure. *J. Pineal Res.* **2022**, p. e12786. <https://doi.org/10.1111/jpi.12786>.
14. Cajochen, C.; Munch, M.; Kobialka, S.; Krauchi, K.; Steiner, R.; Oelhafen, P.; Orgul, S.; Wirz-Justice, A. High sensitivity of human melatonin, alertness, thermoregulation, and heart rate to short wavelength light. *J. Clin. Endocrinol. Metab.* **2005**, *90*, 1311–1316. <https://doi.org/10.1210/jc.2004-0957>.
15. Nagare, R.; Rea, M.S.; Plitnick, B.; Figueiro, M.G. Nocturnal melatonin suppression by adolescents and adults for different levels, spectra, and durations of light exposure. *J. Biol. Rhythm.* **2019**, *34*, 178–194. <https://doi.org/10.1177/0748730419828056>.
16. Allen, A.E.; Hazelhoff, E.M.; Martial, F.P.; Cajochen, C.; Lucas, R.J. Exploiting metamerism to regulate the impact of a visual display on alertness and melatonin suppression independent of visual appearance. *Sleep* **2018**, *41*, zsy100. <https://doi.org/10.1093/sleep/zsy100>.
17. Spitschan, M.; Lazar, R.; Yetik, E.; Cajochen, C. No evidence for an S cone contribution to acute neuroendocrine and alerting responses to light. *Curr. Biol.* **2019**, *29*, R1297–R1298. <https://doi.org/10.1016/j.cub.2019.11.031>.
18. Kennaway, D.J. A critical review of melatonin assays: Past and present. *J. Pineal Res.* **2019**, *67*, e12572. <https://doi.org/10.1111/jpi.12572>.
19. Kennaway, D.J. Measuring melatonin by immunoassay. *J. Pineal Res.* **2020**, *69*, e12657. <https://doi.org/10.1111/jpi.12657>.
20. Chellappa, S.L. Individual differences in light sensitivity affect sleep and circadian rhythms. *Sleep* **2021**, *44*, zsaa214. <https://doi.org/10.1093/sleep/zsaa214>.
21. Spitschan, M.; Santhi, N. Individual differences and diversity in human physiological responses to light. *EBioMedicine* **2021**, *75*, 103640. <https://doi.org/10.1016/j.ebiom.2021.103640>.
22. Phillips, A.J.; Vidafar, P.; Burns, A.C.; McGlashan, E.M.; Anderson, C.; Rajaratnam, S.M.; Lockley, S.W.; Cain, S.W. High sensitivity and interindividual variability in the response of the human circadian system to evening light. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 12019–12024.
23. Spitschan, M.; Lambert, B. melluxdrc: Power Analysis for Human Melatonin Suppression Experiments (Main Package). 2021. Available online: <https://github.com/mellux-project/melluxdrc> (accessed on 20 February 2024).
24. Spitschan, M.; Lambert, B. mellux-app: Power Analysis for Human Melatonin Suppression Experiments (Shiny app). 2021. Available online: <https://github.com/mellux-project/mellux-app> (accessed on 20 February 2024).
25. Wand, M. *KernSmooth: Functions for Kernel Smoothing Supporting Wand & Jones (1995)*, R package version 2.23-15; 2015.
26. Carpenter, B.; Gelman, A.; Hoffman, M.D.; Lee, D.; Goodrich, B.; Betancourt, M.; Brubaker, M.; Guo, J.; Li, P.; Riddell, A. Stan: A probabilistic programming language. *J. Stat. Softw.* **2017**, *76*, 1–32.

27. Vehtari, A.; Gelman, A.; Simpson, D.; Carpenter, B.; Bürkner, P.C. Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC. *arXiv* **2019**, arXiv:1903.08008 .
28. Lambert, B. *A Student's Guide to Bayesian Statistics*; Sage: Newcastle upon Tyne, UK, 2018.
29. Spitschan, M.; Lambert, B. mellux-model_fitting: Power Analysis for Human Melatonin Suppression Experiments (Model Fitting Code). 2021. Available online: <https://www.biorxiv.org/content/10.1101/2023.01.30.526303v1> (accessed on 20 February 2024).
30. Lambert, B. An Introduction to Inverse Transform Sampling. Youtube. 2018. Available online: <https://www.youtube.com/watch?v=rnBbYsysPaU> (accessed on 20 February 2024).
31. CIE S 026:2018; CIE System for Metrology of Optical Radiation for ipRGC-Influenced Responses to Light. CIE: Vienna, Austria, 2018.
32. Spitschan, M.; Mead, J.; Roos, C.; Lowis, C.; Griffiths, B.; Mucur, P.; Herf, M.; Nam, S.; Veitch, J.A. luox: Validated reference open-access and open-source web platform for calculating and sharing physiologically relevant quantities for light and lighting. *Wellcome Open Res* **2022**, 6, 69. <https://doi.org/10.12688/wellcomeopenres.16595.3>.
33. Spitschan, M.; Lambert, B. mellux-Simulations: Power Analysis for Human Melatonin Suppression Experiments (Simulations). 2021. Available online: <https://github.com/mellux-project/mellux-simulations> (accessed on 20 February 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.