COVID-19 strikes again: the effect of increased exposure to electronic screens on Dry Eyes Disease

Lauke Stoel - 6899544

[l.a.stoel@uu.nl](mailto:l.a.stoel@uu.nl)

# Introduction

The aim of this analysis is to investigate whether the severity of Dry Eye Disease (DED) can be explained by the number of hours a student is exposed to electronic screens. This question has become of particular interest since the start of the COVID-19 pandemic, as students are receiving their education in an online format, which plausibly results in more exposure to electronic screens.

The measure used to capture the severity of DED is the Ocular Surface Disease Index (OSDI). The OSDI is a questionnaire designed to provide a rapid assessment of the symptoms of ocular irritation consistent with DED on vision-related functioning (Schiffman, 2000). The OSDI is measured on a scale ranging from 0 to 100, where higher scores represent greater disability. The main predictor of interest is the total amount of time a student spent in front of a screen. This is measured in hours.

The data used for this analysis were collected by means of a survey, administered to 97 Mexican university students (Santa Cruz-Pavlovich et al., 2021). The survey was administered four times over the course of six weeks between April 9th and March 21st of 2020. In week zero, a baseline measurement was taken of the OSDI score and the total number of hours a student spent in front of a screen over the course of three days. The students did not have any classes that week. The subsequent measurements were taken in week two, four and six of the semester. Each contains an OSDI score and the total number of hours a student spent in front of a screen over the course of the same three days, broken down into hours spent on online classes, homework and leisure respectively. For this analysis I aggregate these categories and use the total number of hours spent exposed to digital screens as a predictor. The dataset also contains general information of the study's subjects, namely their subject number, age (in years) and sex (male or female).

The research question I aim to answer is twofold, namely whether the severity of DED in students progresses over the course of the semester and whether that progression can be explained by the number of hours a student is exposed to electronic screens. For simplicity’s sake, I focus on the latter part of the question and conduct the analysis by regressing the difference between students’ OSDI scores in week 6 of the semester and their baseline OSDI score on the difference between their total screen time in that same week and their baseline screen time. This means my analysis only explains variance between individuals, not within individuals, based on difference in screen time between students. The regression analysis includes a parameter controlling for the age of the student, transformed to be grand mean centered (*Mean* = 20.5).

The analysis is performed within the Bayesian framework, using a Gibbs-sampler with a Metropolis-Hastings step. Convergence of the model is assessed using trace plots, the Gelman-Rubin statistic, MC errors, autocorrelations and – in the case of the Metropolis-Hastings step – the acceptance ratio. To judge whether my data meet the assumptions of the multiple linear regression model, skewness and kurtosis of the data are assessed by means of a posterior predictive check (PPC). Model selection is based on the Deviance Information Criterion (DIC) and the hypothesis regarding the effect of screen time on OSDI score is assessed using the Bayes Factor (BF). Parameter estimates are presented as the Expected A Posteriori (EAP) of each parameter with corresponding 95% Central Credible Intervals (CCI).

# Methods

This analysis is conducted using Bayesian methods, in which the data from the sample are combined with my prior belief about the parameters. The parameters estimated are the intercept, the regression coefficients for the predictors age and difference in screen time, and the variance. I use uninformative normal priors for the intercept and the regression coefficient for age with mean and precision . Assuming these variables are normally distributed, this choice of prior is conjugate with the likelihood of the data, resulting in a normal posterior distribution. The uninformative, conjugate prior used for the variance is an Inverse Gamma distribution with shape and rate parameters and , which results in an Inverse Gamma posterior distribution.

The informative prior used for the regression coefficient of difference in screen time is a t-distribution with and . This prior is based on previous research that found that extended screen exposure is associated with increased risk of symptomatic DED compared to no symptomatic DED (Inomata et al., 2020). I elect to use this t-distribution to account for some uncertainty around the estimate of the effect. The use of this t-distribution as a prior yields no known shape of the posterior distribution of this parameter, but we do know the shape of the distribution up to a proportionality constant, warranting the use of Metropolis-Hastings estimation. As a proposal distribution, I use a normal distribution with and , to capture as much of the posterior distribution as possible based on where I expect it to lie.

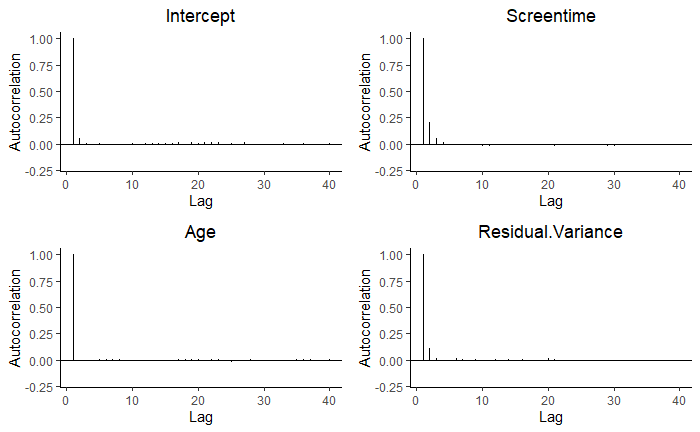
The sampler runs two separate times, resulting in two chains. In the case of the intercept and the two regression coefficients, the starting values for these chains were chosen by picking random values between 0 and 10, based on plausible values on the scale of the OSDI variable. I made sure that each variable’s starting value had the opposite sign in the second chain compared to the first, to ensure that both positive and negative values would be sampled. For the variance, larger, positive starting values were chosen, as variance cannot be negative.

The sampler runs a total of 20000 iterations, creating two chains of each of which the first 1000 iterations are removed to account for the burn-in period.

# Results

## Convergence

Judging from the trace plots, it appears the sampler has converged for each of the four parameters, as both chains are almost completely overlapping (see Figure 1). Furthermore, the autocorrelations of all the parameters reduce to near zero within the first ten iterations, indicating that the sampler moves through the parameter space very quickly (see Figure 2). For the difference in screen time, the acceptance rate is 0.78, which means that 78% of the new proposed values drawn from the proposal distribution are accepted into the chain, indicating a good choice of proposal distribution. Furthermore, the MC error is below 5% of the parameter estimates for all parameters except the variance, which indicates minimal error resulting from the fact that we are using an iterative procedure and that we have a sufficiently large sample (see Table 1). Lastly, the Gelman-Rubin statistic is almost exactly equal to 1 for each parameter, which means that the variance between chains is approximately equal to the within-chain variance and that the sampler has converged well (see Table 1).



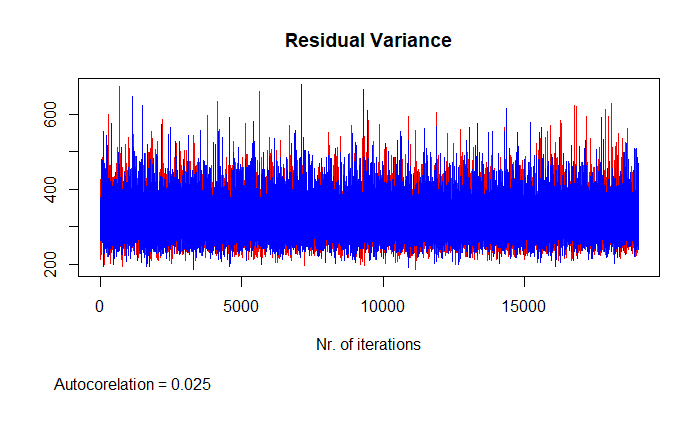
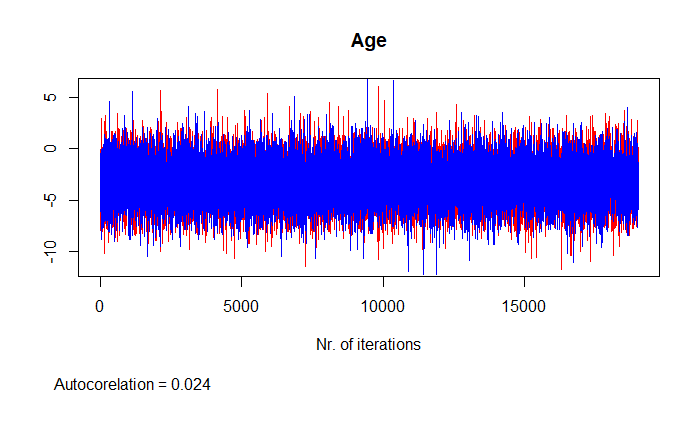
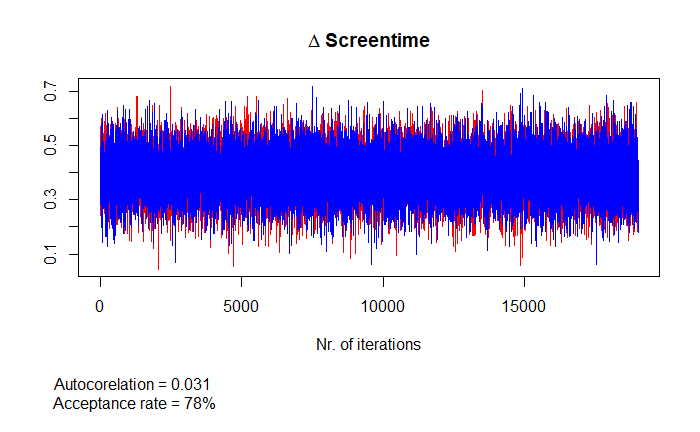
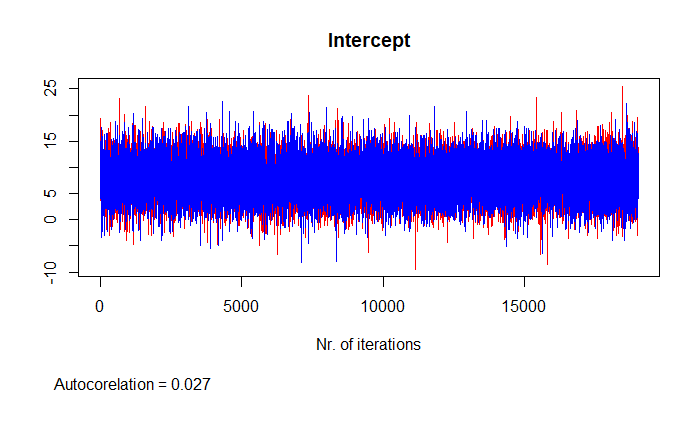


Figure 1: Trace plots of the intercept, variance and the variables age and difference in screen time, using two chains.

Figure 2: Bar plots of the autocorrelation from for each estimated parameters, up to the first 40 iterations of the first chain. The second chain shows similar results.

## Estimates

The EAP estimates are obtained by taking the mean of each parameter over the two chains combined. The model results in Table 1 indicate that there is a positive effect of difference in screen time on the difference in OSDI score. For every hour more a student of average age spent exposed to a screen in week 6 of the semester compared to week 0, the difference in their OSDI score over that same period increases by 0.39 points. The 95% CCI indicates that the probability that the true size of this effect lies between 0.23 and 0.55 is 95%. The fact that the CCI does not encompass zero supports our belief that there is an effect of difference in screen time on difference in OSDI score. The evidence in our data does not indicate a relationship between age and the difference in OSDI score . However, this CCI only barely encompasses zero. Based on previous research, I have reason to believe that age might actually have an effect on DED (Farrand et al., 2017). Interestingly enough, this effect is of a different sign that what is suggested by these parameter estimates. My parameter estimates hint at a negative effect of increased age on severity of DED, while previous studies have concluded that advanced age is associated with increased risk of DED. I further explore the effect of this variable with use of the Bayes Factor under section 3.3.2.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Table 1:** Bayesian linear regression results | | | | | | | |
|  | | **Mean** | **SD** | **MC- error** | **Gelman-Rubin** | **2.5%** | **Median** | **97.5%** |
| Intercept | | 7.897 | 3.409 | 0.031 | 1.000 | 1.169 | 7.915 | 14.539 |
| Screen.diff | | 0.391 | 0.082 | 0.001 | 1.000 | 0.231 | 0.391 | 0.553 |
| Age.gmc | | -3.168 | 1.912 | 0.017 | 1.000 | -6.949 | -3.163 | 0.626 |
| Residual Variance | | 324.080 | 51.723 | 0.472 | 1.000 | 240.104 | 318.703 | 439.842 |

## Model Comparison

### 3.3.1. Using the Deviation Information criterion (DIC)

As we saw from the results in Table 1, the centered predictor age did not have an effect on the difference in OSDI score in week six compared to week zero. It could be of interest to test whether a model with another subject level variable as a second predictor fits the data better than the model using age as a predictor. One way of comparing such models is by means of the DIC. This information criterion weighs the fit and the complexity of the model and also takes the complexity of the parameters into account. The lower the DIC, the more parsimonious the model. The DIC is a comparative measure, meaning that it is equipped to select the best predicting model out of the models under consideration. It does not tell us anything about how well this model performs compared to the ‘true’ model.

An alternative model was fitted with a subject’s sex as a predictor instead of age and a brief check on convergence was performed, with satisfactory results. The DIC of the model including age as a predictor is 817, and the DIC of the model including sex as a predictor is 939. The difference in DIC scores is 82, yielding overwhelming evidence that we should favour the original model over the alternative, as it is more parsimonious.

### 3.3.2. Using the Bayes Factor (BF)

Another way of comparing different models is through the Bayes Factor, which can be used to quantify the support for one hypothesis compared to another. I employ the Bayes Factor to test my earlier findings based on the parameter estimates under section 3.2. I first test the hypothesis that the difference in number of hours spent exposed to a screen between week six and week zero has an effect on the difference in OSDI over that same period. The Bayes Factor for that hypothesis compared to its complement is 633.6, meaning that there is more than 600 times more support for the hypothesis that the effect of difference in screen time is larger than zero than for the hypothesis that it is smaller than or equal to zero.

I also test the hypothesis that advanced age has a positive effect on the difference in OSDI score, as previous research suggests. The Bayes Factor of this hypothesis compared to its complement is 0.018, meaning that – given these data – there is over 50 times more support for the hypothesis that the effect of age is zero or negative, than for the hypothesis that it is positive. This is in direct contradiction with previous research and my own expectation. One explanation for this result could be that this sample only contains students between the ages of 18 and 24 and that the aforementioned effect of older age on DED does not appear until a later age.

In the introduction I mentioned that I would focus on the question whether the difference in OSDI score between week six and week zero could be explained by difference in screen time. With respect to the first part of my research question, I assumed that the symptoms of DED get progressively worse over the course of the semester. The Bayes Factor allows me to test if this assumption was valid. To that end, I test the informative hypothesis that the OSDI score steadily increases over the course of the semester against two other hypotheses, namely that there is only a sharp increase in OSDI score after the second week of the semester (when students actually start working) and the hypothesis that there is no discernible difference in OSDI scores over the semester. The hypotheses are expressed as follows:

The Bayes Factor of the first hypothesis is 2.7, indicating a little less than three times more support for this hypothesis compared to the unconstrained hypothesis. The Bayes Factor of and are 27.3 and 22.5 respectively, indicating great support for these hypotheses over the unconstrained one. This means that out of the three hypotheses, there is the most in these data support for . Nonetheless, it should be noted that this means that hypothesis fits the data best out of all models under consideration; there may be another model that we did not consider, that fits the data better.

## Posterior Predictive Check

To see whether my data actually meet the assumptions of my selected model, I check normality by assessing the skewness and kurtosis of my outcome variable by means of a posterior predictive check. As a proxy for measuring skewness of the data, I created my own test statistic, which counts the number of unique observations higher and lower than the mean, computes their ratio and takes the absolute difference of that ratio from one. This is an indication of how skewed the data is in either direction. This test statistic is calculated for the observed data and for 1000 simulated data sets . These data sets are simulated under the normal model, using 1000 sets of parameter estimates that are randomly sampled from their joint posterior distribution. To obtain said joint posterior distribution, the analysis was run again, but this time with all uninformative priors. The probability that my data actually come from a normal distribution can be approximated by the proportion of test statistics of the simulated data sets that is higher than that of my observed data. In other words, the probability that the skewness of my data falls within the bounds of the skew that can be expected from randomly drawn datasets under my specified model. I will consider any p-value lower than 0.4 as a significant deviation of my data from the normal model. The Bayesian p-value is calculated as follows:

The Bayesian p-value for my own skewness test returns 0.420, which means it is quite likely that data simulated from a normal model display a larger skew in any direction than my data do. This means that based on this test statistic, there is no evidence that my data do not conform with the assumptions of a normal model.

However, this test statistic only tells us something about how many *unique* observations there are higher and lower than the mean of the data set. It does not say anything about how many observations are in each bin, which is still important in determining the shape of the distribution. To that end, I also calculate a Bayesian p-value for testing kurtosis, but this time with the *kurtosis()* function from the *moments* package in R. This p-value returns 0.004, indicating that it is nearly impossible that my data conform to the shape of a normal distribution. Since one of my two elected test statistics failed, I have to conclude that the assumption of normality of my data is not met and that the outcomes of my linear regression analysis are most likely biased.

## Discussion and multi-level extension

As mentioned in the introduction, the analysis performed here is a very simple one and not at all adequate to answer my full research question. For this type of longitudinal data, a different type of analysis would have been more suitable, for example a Growth Curve Model (GCM). The advantage of using a multilevel approach like this is that this analysis can capture between subject variance as well as within subject variance; in this case, the variance in OSDI scores of the same individual over the different measurement occasions. Another advantage is that the measurement occasions need not be equally spaced or contain the same number of observations, although in this data set neither of these situations apply.

One of the advantages of using a Bayesian GCM over a classical one is that it allows us to incorporate our prior knowledge about the effects of exposure to digital screens on DED into our analysis, like I did in the present analysis. It would also allow us to weight different sources of prior information into a power prior, which could be of particular interest for the age parameter in this analysis, as previous research finds an effect of the opposite direction from the one my data hint at. As noted in the results section, this dataset contains only observations from a very homogenous population in terms of age. Weighting the information in previous studies according to their similarity to this study in terms of set up and the sampled population could provide us with a lot more insight into the actual effect that age may have on DED. Another advantage is that in Bayesian GCM, we can quantify the uncertainty around each parameter by means of their posterior distribution, also for the person-specific intercept and slope. Moreover, it has been shown that Bayesian GCM generally yields more accurate measures of uncertainty around the estimates, because it takes into account that the latent, person-specific constructs, the variation of which we see captured in the population level variance, are themselves measured with uncertainty (Oravecz & Muth, 2018).

## Conclusion

Since my data failed to meet the assumptions necessary for running a linear regression analysis, the results of this analysis cannot be interpreted directly. Failure to meet the necessary assumptions could have to do with the fact that I chose to focus on the difference score between only one measurement occasion and the baseline for the purposes of this assignment. Analysing these data using Bayesian GCM may yield more robust results. Furthermore, the third and last Bayes Factor calculated suggests that the severity of these symptoms increase over the course of the semester, although not linearly. Bayesian GCM would be better equipped to analyse a non-linear relationship. Nonetheless, given previous research and the – albeit flawed – indication in my analysis, I conclude there is reason to believe that increased exposure to digital screens is likely a predictor of severity of DED symptoms. All the more reason to head back to campus as soon as it is safe!

## Bibliography

Farrand, K. F., Fridman, M., Stillman, I. Ö., & Schaumberg, D. A. (2017). Prevalence of Diagnosed Dry Eye Disease in the United States Among Adults Aged 18 Years and Older. *American Journal of Ophthalmology*, *182*, 90–98. https://doi.org/10.1016/j.ajo.2017.06.033

Inomata, T., Iwagami, M., Nakamura, M., Shiang, T., Yoshimura, Y., Fujimoto, K., Okumura, Y., Eguchi, A., Iwata, N., Miura, M., Hori, S., Hiratsuka, Y., Uchino, M., Tsubota, K., Dana, R., & Murakami, A. (2020). Characteristics and Risk Factors Associated With Diagnosed and Undiagnosed Symptomatic Dry Eye Using a Smartphone Application. *JAMA Ophthalmology*, *138*(1), 58. https://doi.org/10.1001/jamaophthalmol.2019.4815

Oravecz, Z., & Muth, C. (2018). Fitting growth curve models in the Bayesian framework. *Psychonomic Bulletin & Review*, *25*(1), 235–255. https://doi.org/10.3758/s13423-017-1281-0

Santa Cruz-Pavlovich, F. J., Bolaños-Chang, A. J., Gonzalez-Gonzalez, J. E., Guzman-Castellanos, J. F., Ledesma-Mijares, R. A., Fuentes-Plata, H., Jimenez-Batalla, C., Betancourt-Valencia, J. A., & Navarro Partida, J. (2021). *Data associated with ‘Online education models during COVID-19 are associated with the development and worsening of dry eye disease’.* [Data set]. 4TU.ResearchData.

Schiffman, R. M. (2000). Reliability and Validity of the Ocular Surface Disease Index. *Archives of Ophthalmology*, *118*(5), 615. https://doi.org/10.1001/archopht.118.5.615