Bayesian Analysis of Randomized Controlled Trials

Julian Bautista, Alex Pavlakis, Advait Rajagopal

November 25, 2017

Abstract

Contents

1 Introduction

Set up the paper.

2 Bayesian Data Analysis

2.1 Model development

2.2 Model estimation

Estimate the posterior distribution conditional on prior knowledge and the observed data.

2.3 Posterior predictive checking

Check the fit of the model and the resulting implications of the posterior parameter estimates.
Repeat

3 Comparison of Bayesian Data Analysis to Other Methods

Write about the advantages of BDA over other formats. How are they similar? How are they different?

4 Reproducibility

Describe the philosophy and importance of reproducibility.

Describe the steps of making research reproducible (e.g. sharing code, sharing data, explicitly specifying model, etc.)

5 Impact of Smartphone App on Eating Behavior

Hildebrandt et all (2017) conducted an experiment to test whether the Noom Monitor, a smartphone application, could augment the effect of in-person therapeutic treatment on binge eating behavior. The treatment, known as *guided self-help treatments based on cognitive-behavior therapy* (CBT-GSH), had been shown in previous research to reduce binge eating behavior by 10-50%. The Noom Monitor application was designed to facilitate CBT-GSH. For this example, we consider two research questions from the experiment:

1. Is CBT-GSH more effective at reducing binge eating behavior when facilitated by the Noom Monitor?

2. Does the effect of the Noom Monitor vary over time?

5.1 Experimental design

66 men and women with Bulimia Nervosa (BN) or Binge Eating Disorder (BED) were randomly assigned into two treatment conditions: CBT-GSH (N= 33) or CBT-GSH +
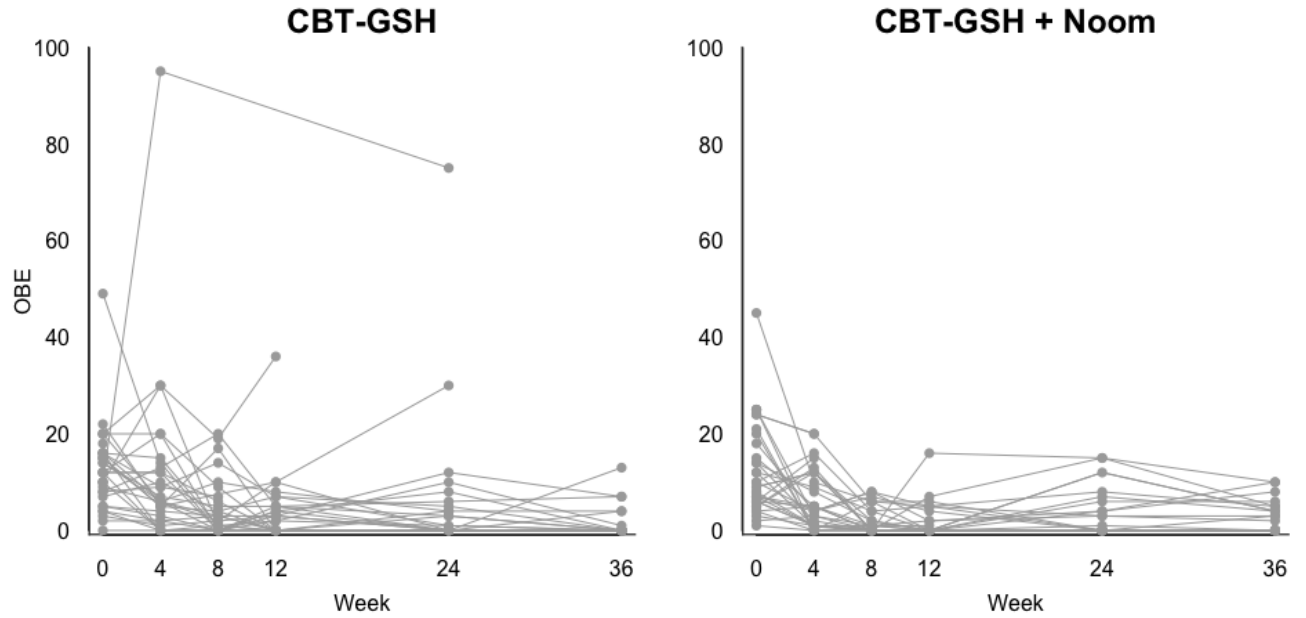
Figure 1: *Plots display OBEs over time for each individual in the CBT-GHS groups (left panel) and CBT-GHS + Noom group (right panel).*

Noom (N=33). Therapy lasted for 12 weeks. Assessments were conducted at weeks 0, 4, 8, 12, 24, and 36. The primary outcome was Objective Bulimic Episodes (OBE).

5.2 Exploratory data analysis

Figure X displays OBEs per week for each individual in both treatment conditions. A few aspects of the data immediately stand out, which suggest that any model should account for individual-level effects and time-level effects, and should let treatment effects vary over time.

- The number of OBEs decreases over the course of the treatment for almost all subjects
- The biggest decreases in OBEs appear to occur in the early stages of treatment
- The primary sources of variation in OBE appear to be *between people* and *over time*.
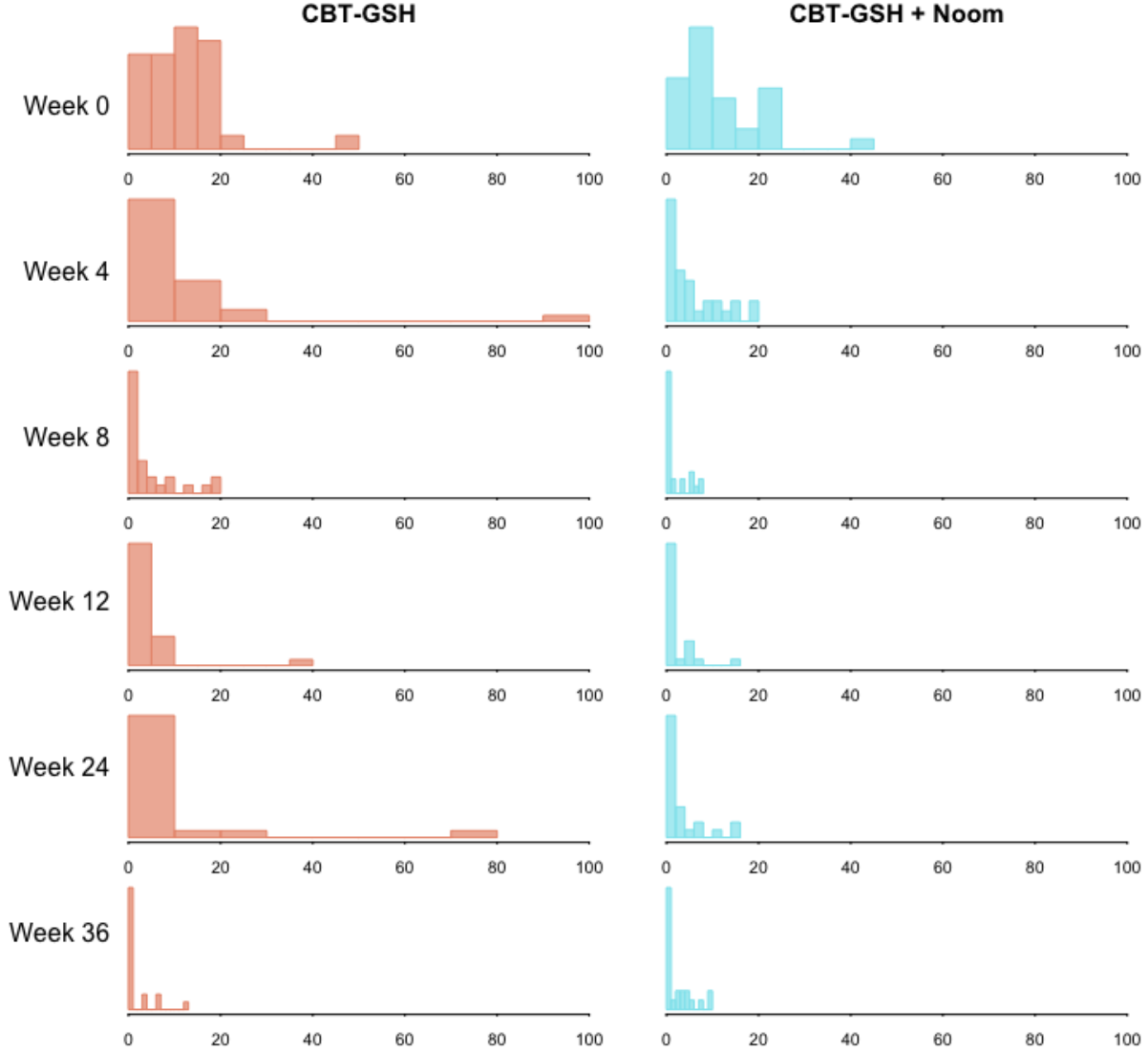
Figure 2: *Histograms display the distribution of OBEs in each condition in each week.*

Figure Y displays the distribution of histograms in each condition in each week. The distributions appear to condense around zero for both conditions over time, and the distributions in the CBT-GSH condition appear to have longer tails than those in the CBB-GSH+Noom condition.

5.3   Model development

We analyze RCTs by modeling the outcome of interest (in this case OBE) as a function of the treatment and all available pre-treatment covariates. The coefficients associated with the treatment reveal the average treatment effects. Inclusion of all available pre-treatment covariates accounts for variation in the outcome variable, decreasing uncertainly around treatment effects and providing the model with more predictive power. We conduct *intent-to-treat* analysis, meaning that our inferences will be based on initial treatment assignment, and will not account for mid-experiment dropouts.

The outcome variable is restricted to be nonnegative integers, so we fit a Poisson regression model, with hierarchies on individuals, time periods, and treatment effects. For each individual in each time period, the number of OBEs follows a Poisson distribution, with a mean dependent on the characteristics of the individual and the time period.

$$OBE_{i,t} \sim Poisson(\lambda_{i,t}) \tag{1}$$

$$\lambda_{i,t} = exp(\alpha_i + \beta_t + \gamma_t T_i + X_i \theta) \tag{2}$$

$$i = 1, ..., 66 \tag{3}$$

$$t = 0, 4, 8, 12, 24, 36 \tag{4}$$

$$T_i = \begin{cases} 0, & \text{if } CBT - GSH \\ 1, & \text{if } CBT - GSH + Noom \end{cases} \tag{5}$$

α is an individual-specific intercept, β is a time-specific intercept, γ is a time-specific treatment effect, T is a treatment indicator, X is a matrix of individual level covariates (age, sex, race, etc), and θ is a vector of individual level effects.

We believe that individual-level intercepts are simultaneously unique to the individual and common to the population; that is, each individual has their own baseline predilection to engage in eating disorder behavior, but their baseline predilections are not drastically different from each other. We operationalize this concept by modeling all individual-level intercepts as coming from a common distribution, with *hyperparameters* μ_α and τ_α.

$$\alpha_i \sim Normal(\mu_\alpha, \tau_\alpha) \; \forall \; i \in 1, ..., 66 \tag{6}$$

Similarly, we believe that time-specific treatment effects may be unique to each period but similar over time. We operationalize this concept by modeling all time-specific treatment effects γ as coming from a common distribution, with *hyperparameters* μ_γ and τ_γ.

$$\gamma_t \sim Normal(\mu_\gamma, \tau_\gamma) \; \forall \; t \in 0, 4, 8, 12, 24, 36 \tag{7}$$

μ_γ is the *grand mean*, the overall treatment effect; τ_γ is the variation in treatment effects over time; and each individual γ_t is a time-period specific treatment effect. This approach has a natural smoothing effect: any extreme estimates of γ_t will be partially-pooled back toward the grand mean μ_γ.

To complete the model we must assign prior distributions to parameters (such as θ) and hyperparameters (such as μ). These prior distributions will serve three main purposes:

1. Transparency: priors allow us to be clear about our modeling assumptions.

2. Convergence: priors help the model "know where to look" and converge.

3. Inclusion of Additional Information: We usually have more information about an experiment than is represented by the data. Priors allow us to include that information explicitly in our model.

For this model, we assign the following prior and hyperprior distributions:

$$\mu_\alpha \sim Normal(5,5) \tag{8}$$
$$\tau_\alpha \sim Cauchy^+(0,30) \tag{9}$$
$$\mu_\gamma \sim Normal(0,5) \tag{10}$$
$$\tau_\gamma \sim Cauchy^+(0,30) \tag{11}$$
$$\theta \sim Normal(0,1) \tag{12}$$
$$\tag{13}$$

The normal distributions around the individual and treatment effects allow us to guide the model to the appropriate range of parameter values, but with wide enough variance (5 in each case) to let the model find its own way in that range. Half cauchy priors on the variance parameters are weakly informative, with much of their mass around zero but gentle slopes in their tails, which have been shown to be effective prior distributions for variance parameters (Gelman, 2006).

5.4   Model estimation

We estimate this model with *Hamilton Monte-Carlo* in Stan. Model code is appended to this document. We find that the model converges with four chains of 2000 iterations each (see table M).

5.5   Posterior predictive checking

Before using our model to make inferences about time-specific treatment effects, we check its fit by comparing model-simulated OBE to data OBE. If model simulations do not track the data well, we may want to revisit our model's assumptions before trusting its inferences. If the model's simulations recover patterns in the data, we are more inclined to trust its inferences.

Figure Q displays OBEs in each period for each individual in each treatment group, from raw data (upper plots) and model simulations (lowers plots). Black lines display means for each. This suggests that the model is broadly able to pick up on the key variables that determine OBE over time for the duration of this experiment.

Another way to check the fit of the model is by comparing simulated data directly against the raw data. Figure P shows this for both treatment conditions. Simulated data for the Noom condition appears to better track the raw data than simulated data for the no Noom condition. This is unsurprising, since the no Noom condition tended to have more outliers, which we would not expect (or want) our model to pick up perfectly from such a small sample.

We conclude our model evaluations by mapping modeled density curves for each condition in each time period over the histograms in figure Y. Figure BB shows that our model is able to broadly pick up on the patterns in the data over time and between treatment conditions.

5.6   Results

Model results are displayed in table M. Results suggest that using the Noom Monitor smartphone application during CBT-GSH may slightly decrease OBEs. There some evidence that
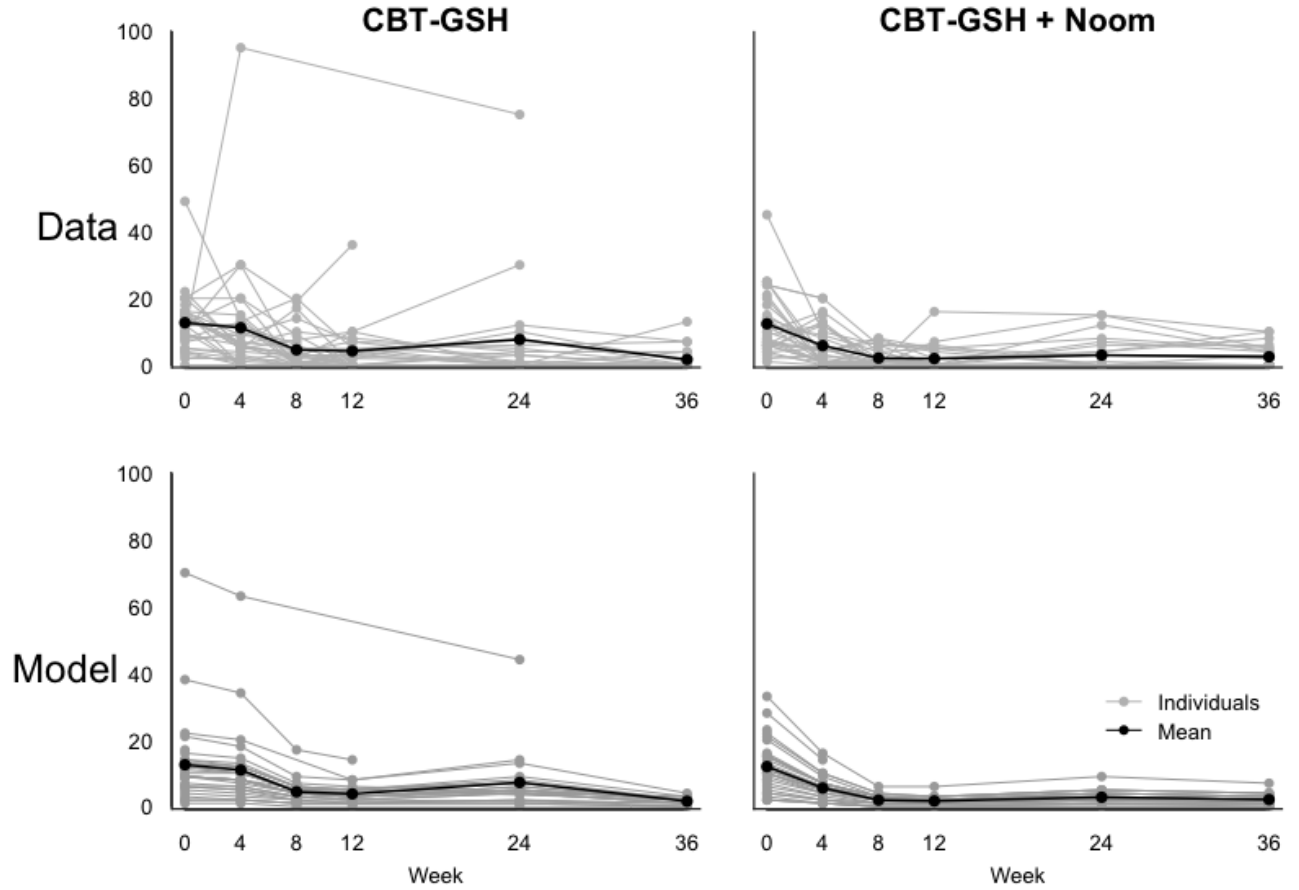
Figure 3: *Upper plots show OBEs in each period for each individual in both treatment group, with black lines representing means. Lower plots show modeled OBEs for each individual in each treatment group, with black lines representing modeled means. The model appears to be able to recover OBEs over time fairly well for both treatment conditions.*
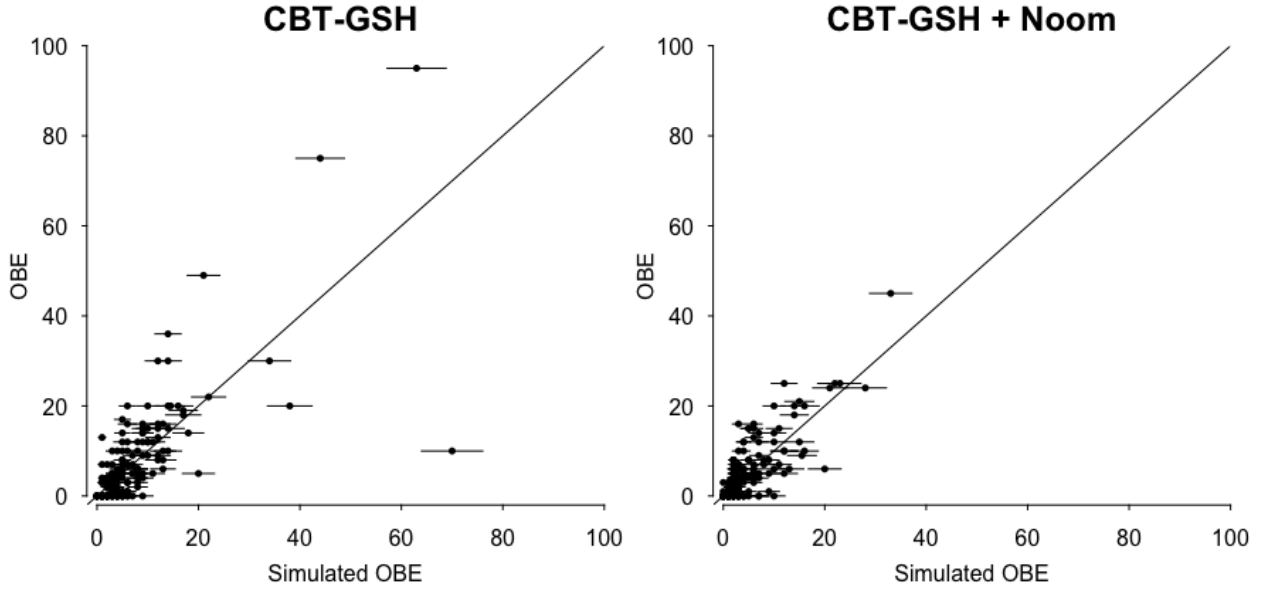
Figure 4: *Plots display siulated OBEs to raw data for both treatment conditions. Model simulations appear to match the raw data well, particularly for the Noom condition, which had fewer outliers.*

the treatment effect varies over time, with the Noom effect being slightly more pronounced during the later stages of therapy.

Figure R displays modeled OBE for both treatment groups (upper plot) and smoothed treatment effects (lower plot). In each measurement period, simulated OBE are higher for the No Noom condition than for the Noom condition, with some of the difference likely attributable to use of the Noom Monitor smartphone app.

6   Conclusion
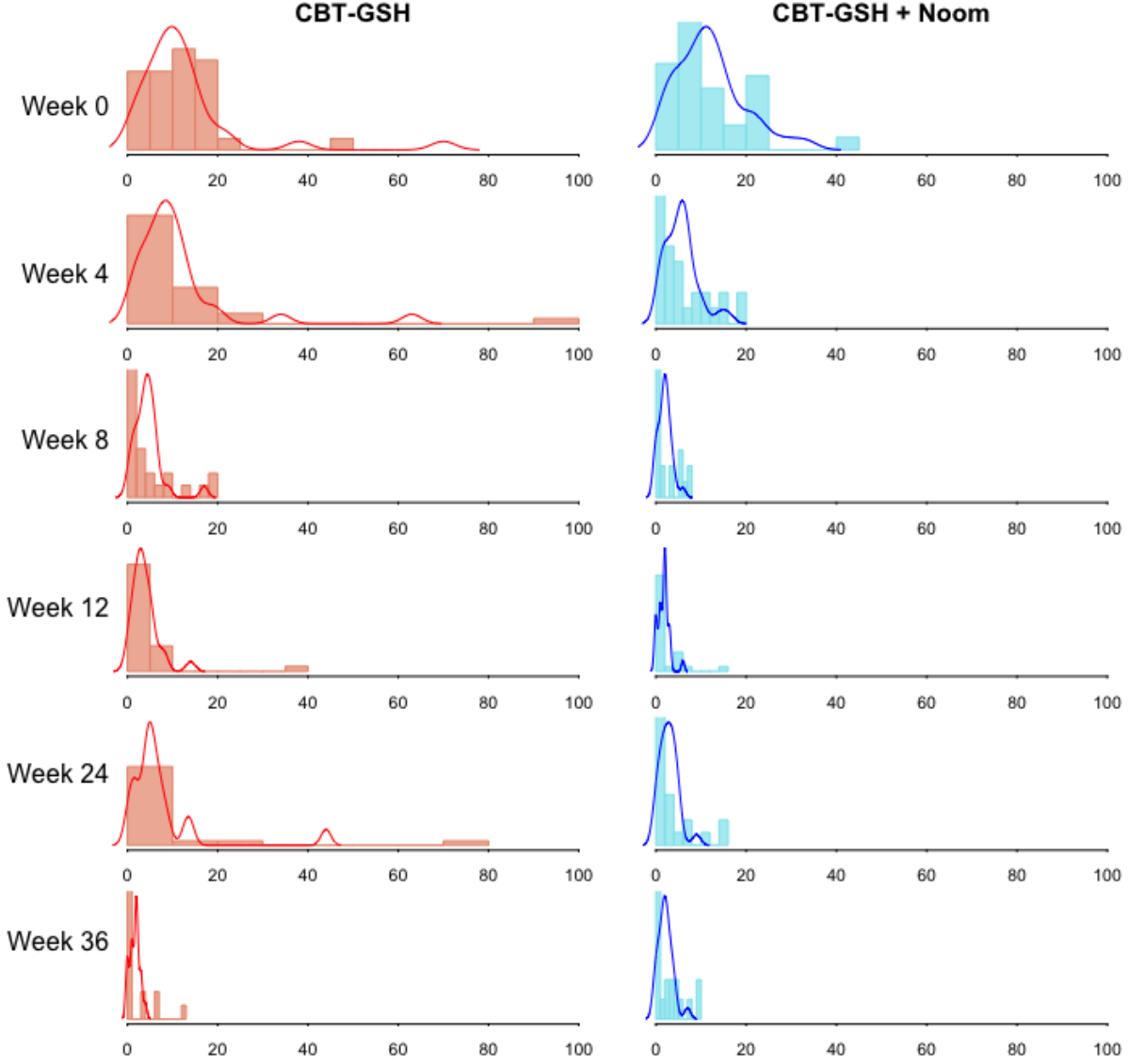
So awesome

7   Appendix

All the math goes here

Figure 5: *Histograms display the distribution of OBEs in each condition in each week.*

	mean	25%	50%	75%
γ_0	0.18	-0.45	0.15	0.78
γ_4	-0.43	-1.05	-0.46	0.16
γ_8	-0.70	-1.33	-0.71	-0.10
γ_{12}	-0.65	-1.28	-0.68	-0.04
γ_{24}	-0.72	-1.34	-0.75	-0.11
γ_{36}	0.21	-0.42	0.19	0.82
μ_γ	-0.34	-0.98	-0.36	0.26
τ_γ	0.64	0.43	0.56	0.77

Table 1: *Table displays model results for Noom effects in all six time periods and grand mean and variance parameters.*
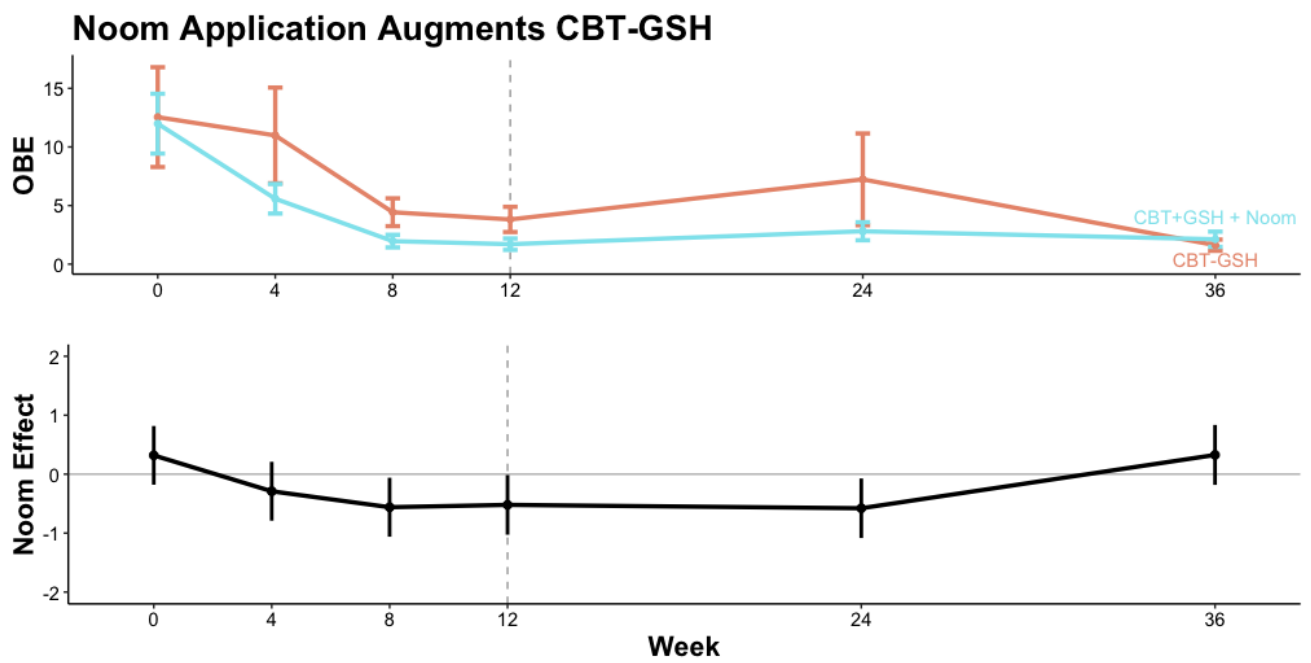
Figure 6: *Upper plot displays modeled OBEs in each time period for the Noom (blue) and no Noom (orange) conditions with 95% intervals. Lower plot displays modeled treatment effects in each period, with 50% intervals.*