**Introduction**

Defining and quantifying uncertainty play vital roles in measuring psychopathology. *Additive models* (AM) maintain that psychopathology is the sum of individual components (e.g., depressed mood = low positive affect + high negative affect [1]) and changes in a linear and additive fashion (e.g., depression at one month = change + depression at baseline). The prevalence of additive models comes from statistical conventions and lack of alternatives, rather than theoretical basis or empirical observations, leading to methodological rigidity that can hinder psychopathology research. For example, research has found treatment-induced change in psychopathology to show multiplicative patterns: Psychotherapy can result in changes proportional to the baseline symptoms wherein post-treatment symptom severity = treatment effect pre-treatment severity [2]–[4]. Thus, we propose *multiplicative models* (MM) that depict psychopathology with greater flexibility and higher faithfulness.

MMs represent psychopathology more faithfully than AMs. AMs imply normality of the psychopathology construct [5]. Normality in psychopathology suggest that those who score low on psychopathology are as rare as those who score high on psychopathology and that such proportions are symmetric. In the general population, however, empirical research shows that data collected on psychopathology follow lognormality. That is, the data are asymmetrically distributed with the majority of the population scoring low on indices of psychopathology and sizably smaller proportion of the population showing high severity [6], [7]. Thus, lognormality from MMs may better describe psychopathology [8], [9].

Conceptually, lognormality in MMs highlights the importance of idiosyncratic characteristics and intra-individual dynamics in assessing severe psychopathology (e.g., [10]). As the variance in lognormal distribution changes given its median, individuals with high severity exhibit highly occasion-varying and person-specific symptoms.

Methodologically, misusing AMs can lead to underpowered studies with as low as 25% statistical power compared to that of MMs. Thus, when correctly specified, MMs can have lower sample size requirement, which can lead to greater cost-efficiency for oftentimes-expensive clinical experiments [11]–[14]. In addition, MMs are invariant across scales, intuitive in interpretation, and robust against outliers [11]. Consequently, MMs have both conceptual and methodological advantages over traditional AMs.

**Method**

The current study focuses on the use of MMs on depression, as measured by PHQ-9 in three panels of the PROMIS open dataset (*n* = 175) [15][[1]](#endnote-1). We examined distributional forms via visualization and the inter-panel and item-rest relations via different AMs and MMs. Inter-panel relations were examined through fixed-effects and mixed-effects linear and polynomial models (see notes in Table 1a and 1b for specific model forms). Item-rest relations were represented via using the individual item to predict the sum of other items (SOI, i.e., sum scores – the individual item score). MMs are applied by first log-transforming all but dummy variables, then fitting respective models, and finally back-transforming certain parameters.[[2]](#endnote-2) This effectively fits log-log fixed-effect and mixed-effect linear models.

**Results**

Lognormality provides a better fit for PHQ-9 scores at all time points than normality [5] (see Figure 1). Consistent with MMs assumptions, the variability in depression scores increases as severity increases (see Figure 2). Both suggest greater plausibility for MMs than AMs.

In modeling depression development across time, multiplicative forms of fixed-effects, fixed-trend (Table 1a), random-intercept, and random-intercept/-slope models (Table 1b) all outperformed their additive counterparts[[3]](#endnote-3). The multiplicative random-intercept random-slope model best captured the development of depression over time. Depression development hinges upon individual levels: those with greater severity can show greater change; as time progresses, the rate of change can also increase. This differs from an additive relationship where a time-invariant change is assumed constant across individuals.[[4]](#endnote-4) MMs accentuate the room for improvement in individuals with high depression severity. Moreover, random-intercept and random-slope components highlight that individuals differ in baseline median[[5]](#endnote-5) and progress at distinct ratios. Altogether, the progression of depression differs across individuals by both random variability and individuals’ prior depression levels in a multiplicative fashion.

Using multilevel models to account for multiple measurement occasions, we report results from AMs and MMs focusing on using individual item to predict SOI (Table 2a and 2b). This shows how the levels of one symptom can relate to the severity in other depressive symptoms. All items showed multiplicative patterns in predicting SOI, encouraging investigation of a multiplicative structure whereby other depression symptoms individual symptom = depression. For example, those who endorse “nearly every day” of “Trouble falling or staying asleep, or sleeping too much” can show depression score 185% higher on other depressive symptoms than those who endorse “Not at all”. Unlike AMs, a ratio effect of 185% in MMs inherently emphasizes individual differences (e.g., for individuals with a baseline of 10, this means 8.5 increase and with a baseline of 5, this means only a 4.25 increase).

**Discussion**

We demonstrated methodological novelty and empirical utility in applying MMs to psychopathology research. The small sample size and the exploratory nature of the study demand future research efforts. Despite these limitations, the study sheds light on the strengths of using MMs to study psychopathology, with regard to both change over time and interconnections among symptoms. MMs more faithfully align with observed data than AMs. MMs offer intuitive interpretation of effects: MMs express effects in terms of ratios (e.g., while an additive effect of 3-point decrease on PHQ-9 loses interpretability to someone unfamiliar with the unit or the range of PHQ-9, a multiplicative effect of 80% is readily interpretable). MMs show greater appreciation of the complexity among symptom connections: the influence of one symptom can vary drastically based on severity of other symptoms. Furthermore, MMs allow methodological freedom, beyond the strict (and sometimes unrealistic) assumptions of AMs, to better depict the empirical observations of and to allow wider theoretical range for psychopathology. Future directions may include formalization of multiplicative measurement models. Altogether, the results of this study suggest that multiplicative models provide a conceptual and methodological stepping stone to better understanding psychopathology constructs and intervention effects.

References:

[1] D. Watson and A. Tellegen, “Toward a consensual structure of mood.,” *Psychol. Bull.*, vol. 98, no. 2, pp. 219–235, 1985.

[2] L. L. Hawley *et al.*, “Cognitive-Behavioral Therapy for Depression Using Mind Over Mood: CBT Skill Use and Differential Symptom Alleviation,” *Behav. Ther.*, vol. 48, no. 1, pp. 29–44, Jan. 2017.

[3] S. Occhipinti, S. K. Chambers, S. Lepore, J. Aitken, and J. Dunn, “A longitudinal study of post-traumatic growth and psychological distress in colorectal cancer survivors,” *PLoS One*, vol. 10, no. 9, pp. 1–12, 2015.

[4] N. L. Kocovski, J. E. Fleming, L. L. Hawley, M.-H. R. Ho, and M. M. Antony, “Mindfulness and acceptance-based group therapy and traditional cognitive behavioral group therapy for social anxiety disorder: Mechanisms of change,” *Behav. Res. Ther.*, vol. 70, pp. 11–22, Jul. 2015.

[5] H. Cramér, “On the composition of elementary errors,” *Scand. Actuar. J.*, vol. 1928, no. 1, pp. 13–74, Jan. 1928.

[6] S. Tomitaka and T. A. Furukawa, “Mathematical model for the distribution of major depressive episode durations,” *BMC Res. Notes*, vol. 7, no. 1, p. 636, 2014.

[7] S. Tomitaka, Y. Kawasaki, K. Ide, H. Yamada, H. Miyake, and T. A. Furukaw, “Distribution of Total Depressive Symptoms Scores and Each Depressive Symptom Item in a Sample of Japanese Employees,” *PLoS One*, vol. 11, no. 1, p. e0147577, Jan. 2016.

[8] R. Gibrat, *Les Inégalités économiques*, 1st ed. Paris, France: Recueil Sirey, 1931.

[9] E. Limpert, W. A. Stahel, and M. Abbt, “Log-normal Distributions across the Sciences: Keys and Clues,” *Bioscience*, vol. 51, no. 5, p. 341, 2001.

[10] M. J. V. Fennell and J. D. Teasdale, “Cognitive therapy for depression: Individual differences and the process of change,” *Cognit. Ther. Res.*, vol. 11, no. 2, pp. 253–271, Apr. 1987.

[11] Q. Liu and S. E. Maxwell, “Multiplicative treatment effects in randomized pretest-posttest experimental designs.,” *Psychol. Methods*, Jun. 2019.

[12] Open Science Collaboration, “Estimating the reproducibility of psychological science,” *Science (80-. ).*, vol. 349, no. 6251, pp. aac4716–aac4716, Aug. 2015.

[13] J. L. Tackett, C. M. Brandes, K. M. King, and K. E. Markon, “Psychology’s Replication Crisis and Clinical Psychological Science,” *Annu. Rev. Clin. Psychol.*, vol. 15, no. 1, pp. 579–604, May 2019.

[14] H. Pashler and C. R. Harris, “Is the Replicability Crisis Overblown? Three Arguments Examined,” *Perspect. Psychol. Sci.*, vol. 7, no. 6, pp. 531–536, Nov. 2012.

[15] P. Pilkonis, “PROMIS 1 Wave 2 Depression.” Harvard Dataverse.

[16] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd Editio. New York: Springer-Verlag, 2002.

*Table 1a. Fixed Effects and Trend Analysis Models for Inter-Panel Relations*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | AIC | Predictors | Estimates1 | SE2 | 95% CI | *t* | p |
| Model 13 | 3359.46 | One Month | -3.69 | .631 | (-4.93, -2.45) | -5.86 | <.001 |
| Three Month | -5.49 | .625 | (-6.71, -4.27) | -8.77 | <.001 |
| Model 23 | 3318.574 | One Month | 82.0% | .033 | (76.8%, 87.5%) | -5.98 | <.001 |
| Three Month | 72.9% | .033 | (68.3%,77.8%) | -9.58 | <.001 |
| Model 33 | 3359.46 | Linear | -49.16 | 5.91 | (-60.76,-37.56) | -8.33 | <.001 |
| Quadratic | 19.65 | 5.91 | (8.05, 31.25) | 3.33 | <.001 |
| Model 43 | 3318.564 | “Linear” | -0.105 | 0.011 | (-0.12,-0.08) | -9.09 | <.001 |

Note 1. Estimates for additive and multiplicative models are respective unstandardized regression coefficient beta estimates and the ratio zeta estimates [11] compared with the baseline.

Note 2. Standard error estimates for multiplicative models are that of the logged model.

Note 3. Let Y, M1, M3, and M respectively represent PHQ-9 sum scores, dummy coded month = 1, dummy coded month = 3, and month as a continuous variable. We have additive fixed effects model 1, ; multiplicative fixed effects model 2, where is the geometric grand mean, where is the ratio effect from month and is the multiplicative error term; additive fixed trend analysis model 3, , here centering is necessary due to multicollinearity between linear and quadratic terms; and multiplicative fixed trend analysis model 4, , here quadratic term is not included because multiplicative pattern may approximate quadratic effects given solely the “linear” term.

Note 4. ABC procedure [11] was employed to obtain a comparable AIC: As multiplicative models are estimated via log-log regression, the AIC from fitted log-log model is not comparable to the AIC value from fitted additive model due to difference in dependent variable. ABC procedure adjusts for this difference and utilizes a modified Box-Cox transformation to render a comparable AIC value for the multiplicative model.

Note 5. At month one, the depression is the (1+1)-.1= 93% of baseline and at month three, is (3+1)-.1=87% of the baseline.

Table 1b. Random-intercept and Random-intercept/-slope Model for Inter-Panel Relations

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | AIC | Fixed Effect Components | | | | |
|  |  | Estimates1 | SE1 | 95% CI | *t* | p |
| Model 52 | 3248.52 | -1.73 | .145 | (-2.01, -1.73) | -11.87 | <.001 |
| Model 62 | 3202.583 | -0.234 | .017 | (-0.26, -0.20) | -13.65 | <.001 |
| Model 72 | 3242.69 | -1.74 | .158 | (-2.05, -1.43) | -11.06 | <.001 |
| Model 82 | 3159.953 | -0.234 | .019 | (-0.27, -0.20) | -12.63 | <.001 |
|  |  | Random Effects Components | | | | |
|  |  | 5 | 5 | 5 | 5 | ICC5 |
| Model 52 | see above | 4.292 | NA | NA | 4.122 | .520 |
| Model 62 | see above | 0.2226 | NA | NA | 0.2186 | .511 |
| Model 72 | see above | 4.340 | 1.202 | -0.092 | 3.583 | .595 |
| Model 82 | see above | 0.2116 | 0.0816 | -0.3136 | 0.1786 | .616 |

Note 1. Estimates for additive and multiplicative models are respective unstandardized regression coefficient beta estimates and the ratio zeta estimates [11] compared with the baseline. Standard error estimates for multiplicative models are that of the logged model.

Note 2. Let Y, M, and respectively represent PHQ-9 sum scores, month, random intercept, and random slope. We have additive random-intercept model 5, ; multiplicative random-intercept model 6, where the random effects are modeled onto the logged scores and is the multiplicative error term; additive random-intercept random-slope model 7, ; and multiplicative additive random-intercept random-slope model 8, where the random effects are modeled onto the logged scores.

Note 3. ABC procedure [11] was employed to obtain comparable AICs. We also considered additive model with log-transformed dependent variable: The random-intercept and random-intercept/-slope model had AICs of 3220.26 and of 3192.68. This implies that the plausibility of MMs does not come solely from the better distributional fit than AMs but from the faithfulness of the inherent multiplicative form to the data.

Note 4. At month one, the depression is the (1+1)-.23 or -.24= 85% of baseline and at month three, is (3+1)-.23 or -.24=73% of the baseline.

Note 5. and respectively denote the standard deviation of the random intercept and random slope, denotes the correlation between random intercept and random slope, is the residual standard deviation, and ICC is the sample intra-class correlation.

Note 6. Estimates for the random components in multiplicative models are with respect to the log-log linear mixed effect models as model 6 can be rewritten as log( and model 8 can be rewritten as log(. Both can be readily estimated with existing software.

*Table 2a. Model Fit of Item-Rest Relations*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Additive | | Multiplicative | | Model Fit Difference |
| item | AICAM | F (3, 327) | AICMM | F (3, 327) |  |
| 1 | 2857.17 | 121.06 | **2819.06** | 101.49 | 38.11 |
| 2 | 2753.57 | 184.18 | **2723.86** | 150.30 | 29.71 |
| 3 | 3022.23 | 46.89 | **2943.72** | 43.33 | 78.51 |
| 4 | 2769.17 | 162.65 | **2679.11** | 157.48 | 90.06 |
| 5 | 3001.44 | 52.06 | **2945.05** | 39.62 | 56.39 |
| 6 | 2841.80 | 118.05 | **2831.46** | 89.16 | 10.34 |
| 7 | 2870.20 | 111.44 | **2849.58** | 85.15 | 20.62 |
| 8 | 3035.88 | 49.42 | **2995.17** | 36.64 | 40.71 |
| 9 | 3102.71 | 34.37 | **3052.23** | 25.22 | 50.48 |

Note 1. All model *p* values are under .001 after multiplicity correction using Bonferroni’s method.

Note 2. *F* values refer to the tests for the effect of the individual items predicting *SOI*s.

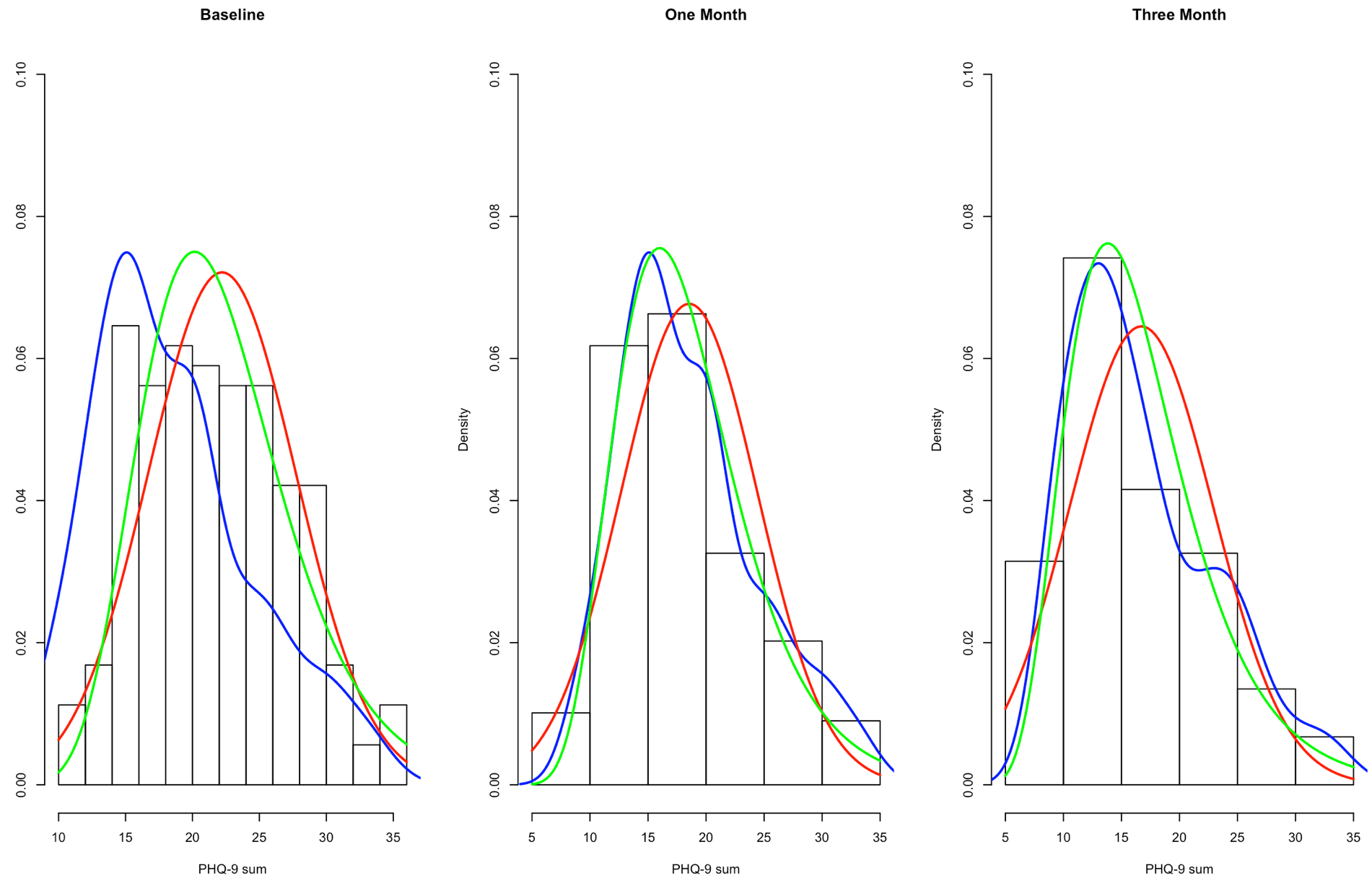
Note 3. AIC values two units smaller than that of additive models are bolded, indicating that multiplicative model is superior in modeling the respective relationship. The smaller the AIC, the better the model fit.

*Table 2b. Multiplicative Models on Item-Rest Relations*

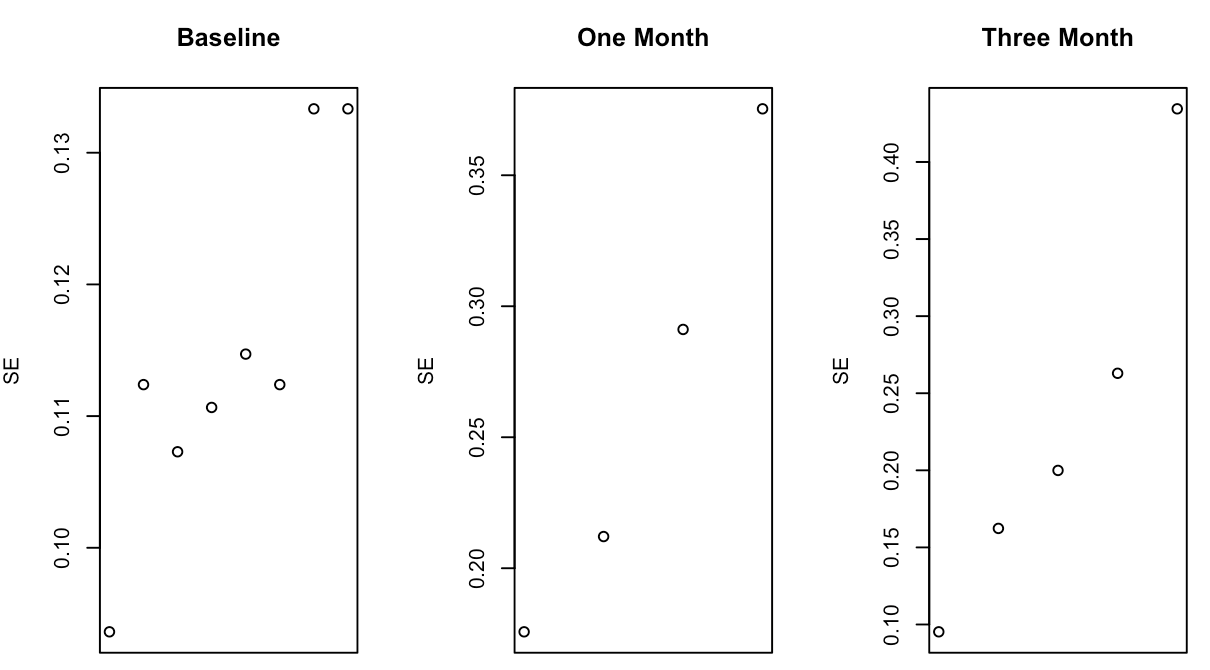
|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Additive Model | | | | | Multiplicative Model | | | | |
|  | score |  | SE | *t* | 95% CI | |  | SE | *t* | 95% CI | |
|  | 1 | 3.11 | 0.427 | 7.27 | 2.27 | 3.95 | 124% | 0.026 | 8.16 | 118% | 131% |
| item 1 | 2 | 6.71 | 0.519 | 12.93 | 5.69 | 7.73 | 151% | 0.032 | 12.86 | 142% | 161% |
|  | 3 | 9.95 | 0.560 | 17.76 | 8.85 | 11.05 | 175% | 0.034 | 16.31 | 163% | 187% |
|  | 1 | 3.44 | 0.395 | 8.70 | 2.66 | 4.21 | 127% | 0.025 | 9.77 | 121% | 134% |
| item 2 | 2 | 8.34 | 0.489 | 17.05 | 7.38 | 9.30 | 166% | 0.031 | 16.57 | 156% | 176% |
|  | 3 | 11.29 | 0.533 | 21.16 | 10.24 | 12.34 | 188% | 0.033 | 19.34 | 176% | 201% |
|  | 1 | 3.44 | 0.395 | 8.70 | 2.66 | 4.21 | 127% | 0.025 | 9.77 | 121% | 134% |
| item 3 | 2 | 8.34 | 0.489 | 17.05 | 7.38 | 9.30 | 166% | 0.031 | 16.57 | 156% | 176% |
|  | 3 | 11.29 | 0.533 | 21.16 | 10.24 | 12.34 | 188% | 0.033 | 19.34 | 176% | 201% |
|  | 1 | 2.39 | 0.480 | 4.99 | 1.45 | 3.34 | 123% | 0.028 | 7.27 | 116% | 130% |
| item 4 | 2 | 6.57 | 0.549 | 11.97 | 5.49 | 7.65 | 158% | 0.033 | 13.86 | 148% | 168% |
|  | 3 | 9.79 | 0.538 | 18.19 | 8.73 | 10.85 | 185% | 0.032 | 18.93 | 173% | 197% |
|  | 1 | 1.73 | 0.454 | 3.80 | 0.83 | 2.62 | 112% | 0.026 | 4.24 | 106% | 118% |
| item 5 | 2 | 4.13 | 0.498 | 8.31 | 3.15 | 5.11 | 126% | 0.029 | 7.86 | 119% | 133% |
|  | 3 | 6.65 | 0.568 | 11.71 | 5.53 | 7.77 | 140% | 0.033 | 10.20 | 131% | 150% |
|  | 1 | 2.77 | 0.379 | 7.30 | 2.02 | 3.51 | 120% | 0.024 | 7.61 | 114% | 126% |
| item 6 | 2 | 6.00 | 0.470 | 12.76 | 5.07 | 6.92 | 142% | 0.029 | 11.93 | 134% | 150% |
|  | 3 | 8.97 | 0.504 | 17.80 | 7.98 | 9.97 | 161% | 0.031 | 15.41 | 152% | 171% |
|  | 1 | 2.50 | 0.397 | 6.30 | 1.72 | 3.28 | 116% | 0.024 | 6.30 | 111% | 122% |
| item 7 | 2 | 5.91 | 0.510 | 11.60 | 4.91 | 6.92 | 140% | 0.030 | 10.93 | 131% | 148% |
|  | 3 | 9.95 | 0.573 | 17.36 | 8.82 | 11.08 | 167% | 0.034 | 15.04 | 156% | 179% |
|  | 1 | 2.90 | 0.468 | 6.20 | 1.98 | 3.82 | 115% | 0.026 | 5.45 | 110% | 121% |
| item 8 | 2 | 5.24 | 0.621 | 8.44 | 4.02 | 6.46 | 130% | 0.034 | 7.70 | 122% | 139% |
|  | 3 | 9.68 | 0.985 | 9.82 | 7.74 | 11.62 | 155% | 0.054 | 8.13 | 139% | 172% |
|  | 1 | 4.75 | 0.579 | 8.20 | 3.61 | 5.89 | 125% | 0.032 | 7.11 | 118% | 133% |
| item 9 | 2 | 7.68 | 1.110 | 6.92 | 5.50 | 9.87 | 143% | 0.060 | 6.02 | 127% | 161% |
|  | 3 | 7.03 | 1.615 | 4.35 | 3.85 | 10.21 | 135% | 0.090 | 3.37 | 113% | 162% |

Note 1. All *p* values are under .001 even after multiplicity correction.

Note 2. Except for item 3, all item-rest relationships were examined using random-intercept random-slope model. For additive models, we have where and are respectively the fixed and random intercept; are the fixed and random slope for the time effect given three-panel data, is the effect of scoring *j* at the individual item on the *SOI.* For multiplicative models, we have where and are respectively the fixed and random intercept; are the fixed and random power parameters for the time effect given three-panel data, is the ratio effect of scoring *j* at the individual item on the *SOI.* For item 3, the were omitted from both additive and multiplicative models due to lack of convergence, resulting in random-intercept models.



*Figure 1. Observed Distributions of PHQ-9 Sum Scores at Baseline, Month one, and Month three*. The blue line represents the empirical density, the red line is the normal density reference line, and the green line shows the lognormal density reference line. The distribution of the reference lines closest to the empirical density line is more likely to be the underlying distribution. As shown here, lognormal distribution provides better fit than normal distribution at baseline, month one, and month three.



*Figure 2. Standard Errors of Binned PHQ-9 sum scores*. PHQ-9 sum scores are first binned equidistantly. Standard errors within each bin is calculated as the division between the variance of and the count of scores in the respective bin. Standard errors, as an indicator of variabilities within respective score ranges, should follow a horizontal line were the underlying model normal. As shown here, the increasing pattern suggest that multiplicative models with lognormal distribution provide better fit for the data.

Notes

1. It is worth noting that multiplicative models can only apply to positive-valued data: We adjust the each individual item on PHQ-9 to score one to four and coded baseline, month 1, and month 3 as 1, 3, and 4. This does not change the model fit or estimates for additive models as the model fit of additive models is invariant to linear transformation. [↑](#endnote-ref-1)
2. The current version of R package `lamme` on CRAN provides straightforward tool for fixed effect models. Functionalities for linear mixed effect models are under development. [↑](#endnote-ref-2)
3. This is based on AIC: Lower AIC indicates better-fitting models, difference in AIC greater than two means sufficiently less support from the data for the model with higher AIC [16] [↑](#endnote-ref-3)
4. This also differs from baseline change interaction effects, as models with interaction effects can show low parsimony and difficult inference: The interaction effects still result in linear changes where the difference in change per difference in baseline is constant [11]. However, despite the consistent ratio effect estimate in multiplicative models, the difference in change increases as the difference in baseline becomes larger. [↑](#endnote-ref-4)
5. Multiplicative models, instead of modeling the arithmetic expected value of the outcome, centers upon the geometric expected value of the outcome. Given lognormality, the geometric mean is equivalent to median. This offers multiplicative greater robustness against outliers, compared with additive models.

   **Contribution**: I conceived and designed the project, identified and retrieved the related data, and completed and interpreted the analysis. All data and scripts will be available on Github by the time of the conference. [↑](#endnote-ref-5)