
What is the numerical nature of pain relief?

Andrew D. Vigotsky¹, Siddharth Tiwari², James W. Griffith³, A V. Apkarian^{3*}

¹Northwestern University, United States, ²Illinois Mathematics and Science Academy, United States, ³Feinberg School of Medicine, Northwestern University, United States

Submitted to Journal:
Frontiers in Pain Research

Specialty Section:
Pain Research Methods

ISSN:
2673-561X

Article type:
Hypothesis and Theory Article

Received on:
10 Aug 2021

Accepted on:
21 Sep 2021

Provisional PDF published on:
21 Sep 2021

Frontiers website link:
www.frontiersin.org

Citation:
Vigotsky AD, Tiwari S, Griffith JW and Apkarian AV(2021) What is the numerical nature of pain relief?. *Front. Pain Res.* 2:756680. doi:10.3389/fpain.2021.756680

Copyright statement:

© 2021 Vigotsky, Tiwari, Griffith and Apkarian. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

This Provisional PDF corresponds to the article as it appeared upon acceptance, after peer-review. Fully formatted PDF and full text (HTML) versions will be made available soon.

What is the numerical nature of *pain relief*?

Andrew D. Vigotsky¹, Siddharth Tiwari², James W. Griffith³, and A. Vania Apkarian⁴

¹Departments of Biomedical Engineering and Statistics, Northwestern University, Evanston, IL

²Illinois Mathematics and Science Academy, Aurora, IL

³Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL

⁴Center for Translational Pain Research and Departments of Physiology, Anesthesia, and Physical Medicine & Rehabilitation, Feinberg School of Medicine, Northwestern University, Chicago, IL

Abstract

Pain relief, or a decrease in self-reported pain intensity, is frequently the primary outcome of pain clinical trials. Investigators commonly report pain relief in one of two ways: using raw units (additive) or using percentage units (multiplicative). However, additive and multiplicative scales have different assumptions and are incompatible with one-another. In this work, we describe the assumptions and corollaries of additive and multiplicative models of pain relief to illuminate the issue from statistical and clinical perspectives. First, we explain the math underlying each model and illustrate these points using simulations, for which readers are assumed to have an understanding of linear regression. Next, we connect this math to clinical interpretations, stressing the importance of statistical models that accurately represent the underlying data; for example, how using percent pain relief can mislead clinicians if the data are actually additive. These theoretical discussions are supported by empirical data from four longitudinal studies of patients with subacute and chronic pain. Finally, we discuss self-reported pain intensity as a measurement construct, including its philosophical limitations and how clinical pain differs from acute pain measured during psychophysics experiments. This work has broad implications for clinical pain research, ranging from statistical modeling of trial data to the use of minimal clinically important differences and patient-clinician communication.

1 Introduction

Pain is highly prevalent, burdensome, and a common reason for doctor visits [Gaskin and Richard, 2012, St Sauver et al., 2013, Vos et al., 2020, Mntyselk et al., 2001]. In an attempt to understand the severity of the patients' pain, doctors and researchers ask patients about the intensity of their pain, requiring patients to condense and transmute their subjective experience to a single number. Despite its abstract and reductionist nature, self-reports of pain intensity are moderately-to-strongly correlated with several patient-reported outcome variables, including quality of life, disability, and more [Yazdi-Ravandi et al., 2013, Nasution et al., 2018]. Moreover, self-reports of pain intensity are remarkably easy and inexpensive to collect. These pragmatic and measurement properties make a reduction in self-reported pain, which we define as pain relief, the gold standard for assessing pain improvement.

Clinical studies of pain commonly quantify pain relief as the primary outcome. However, how pain relief is quantified and reported roughly falls into one of two categories: absolute reductions in pain and relative (or percent) reductions in pain. For example, studies that report absolute reduction may state that a drug decreased pain by 2/10 numerical rating scale (NRS) units or 23/100 visual analog scale (VAS) units. Alternatively, studies that report relative reductions may state that pain decreased by 13 percentage units more in the drug group relative to the placebo group. Although both approaches to reporting pain reductions are common, they are conceptually incompatible (unless baseline pain is perfectly homogeneous; see *Statistical Background*). Their incompatibility begs the question as to whether one approach is more appropriate than the other.

In this paper, we aim to illuminate the issue of absolute versus relative pain relief.¹ We rely on statistical theory to provide researchers and statistically-minded clinicians with the background necessary to understand

¹For simplicity, herein, we will refer to self-reported pain intensity simply as pain.

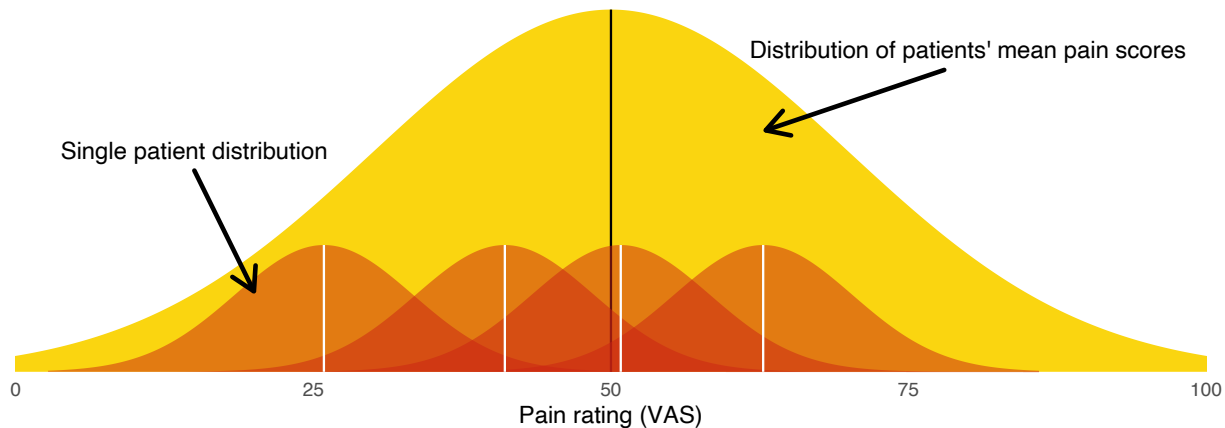


Figure 1: Graphical illustration of the hierarchical model from which patients’ pain scores are sampled. The broad yellow (light grey) distribution is the between-patient distribution (level 2), from which each patient’s mean pain score is sampled. Each red (dark grey) distribution is a within-patient distribution (level 1), from which single measurements are sampled.

45 these measurement models, for which readers are assumed to be familiar with linear regression. In addition,
 46 we empirically analyze four datasets to reinforce and make tangible our conceptual discussion.

47 2 Statistical Background

48 Whenever one uses data to make a calculation, they are building a model. Every model has assumptions,
 49 but still, models should accurately reflect the data they are intending to simplify and thus represent. With
 50 regards to modeling pain relief, when reporting absolute changes in pain, one is assuming the process is
 51 additive. Alternatively, when reporting percent changes in pain, one is assuming the process is multiplicative.
 52 These assumptions have corollaries that *prima facie* may be unclear. In this section, we aim to explain the
 53 processes that would generate each of these models and the theoretical implications of these measurement
 54 and modeling assumptions.

55 2.1 Additive Model

56 The additive model and its implications are best understood by defining a *data-generating process*. This
 57 involves creating a mathematical model that reflects how one thinks the data are created. Because longi-
 58 tudinal pain relief is of interest, there is commonly at least one pain rating at the beginning of the study
 59 (x_i) and at least one or more follow-up ratings (y_i) for each subject i . The additive model of pain relief
 60 uses the simple difference between these pain ratings to calculate absolute pain relief ($\delta_i = y_i - x_i$), where
 61 negative δ_i ’s indicate relief and positive δ_i ’s indicate worsening of pain. Although straightforward, this is a
 62 gross oversimplification.

63 In reality, the pain data are messy. For one, between-patient heterogeneity is appreciable—pain ratings at
 64 intake will often range from the minimum required for study entry (e.g., 4/10 NRS) to the scale’s maximum
 65 (e.g., 10/10 NRS). In addition, patients’ pain fluctuates from minute-to-minute, hour-to-hour, day-to-day,
 66 and so on. To complicate matters further, the process of converting a qualia to a number is undoubtedly
 67 fuzzy, meaning the pain ratings themselves will have noise associated with them. Thus, there are two sources
 68 of variance to consider: between patients and within patients. These sources of variance can be thought of
 69 hierarchically (Figure 1).

70 Between-patient heterogeneity is a natural place to start. The entire sample of patients will have a mean
 71 pain score μ . Each patient’s mean at baseline, α_i , will be dispersed around this group mean according to

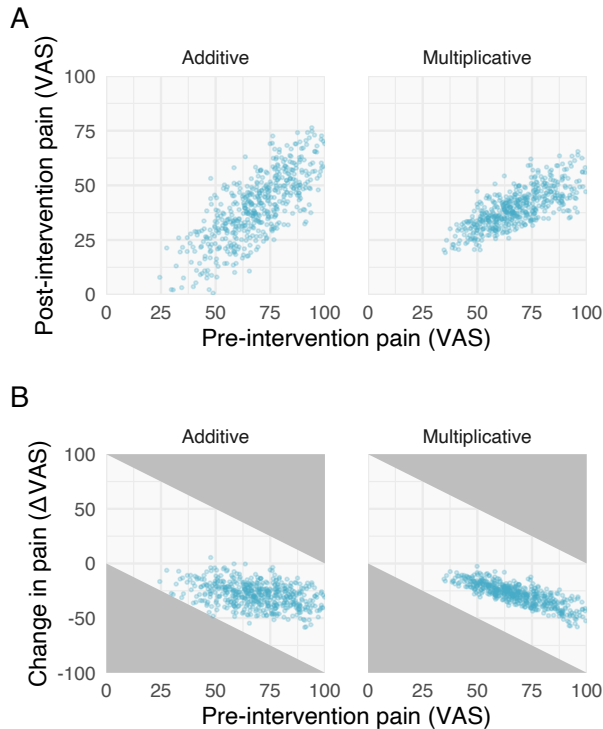


Figure 2: Properties of additive and multiplicative data. We simulated data with additive (top) and multiplicative (top) assumptions. **(A)** Relationships between pre- and post-intervention pain scores when improvements are additive (top) and multiplicative (bottom). Note the additive post-intervention scores are relatively homoscedastic, while the variance of multiplicative post-intervention scores increases with increasing pre-intervention scores. **(B)** Negative relationships between change scores and pre-intervention scores. Grey areas in **(B)** represent regions where points are not possible due to measurement constraints; that is, because a change score cannot be $> |100|$.

72 the between-subject variance τ^2 . We can say that patient means are distributed

$$\alpha_i \sim \mathcal{N}(\mu, \tau^2).$$

73 This distribution of patient means is illustrated in yellow in Figure 1.

74 The notion of within-patient heterogeneity implies there will be variance around each patient’s mean pain.
 75 When we ‘sample’ a patient’s pain rating, we do not observe α_i ; rather, we obtain a value $\alpha_i \pm \sigma$. These
 76 within-patient distributions are illustrated in red in Figure 1. Together, the within- and between-patient
 77 models form a hierarchical model (Appendix A).

78 Because the patient’s pre- and post-intervention pain ratings have variability associated with them, the
 79 observed difference scores are subject to regression toward the mean (RTM). RTM is a statistical phenomenon
 80 whereby higher initial scores are likely to be followed by lower measurements, and similarly, lower initial
 81 scores are likely to be followed by higher measurements. For example, suppose someone’s diastolic blood pressure
 82 is normally around 70 mmHg. If a doctor measures that individual’s blood pressure and finds it to be 90
 83 mmHg, it is highly probable that the next time it is measured, it will be lower than 90 mmHg. Individuals
 84 whose measurements deviate more from their mean will thus appear to undergo greater changes. In the
 85 case of a pain study, those who start off with greater pain levels will regress toward the mean, in turn
 86 creating larger change scores. Importantly, this phenomenon is purely statistical and can be explained by
 87 the reliability of the measurement.

88 Measurement reliability is commonly quantified using the intraclass correlation coefficient (ICC). The
 89 simplest version of the ICC is the ratio of the between-patient variance to the total variance,

$$\frac{\tau^2}{\tau^2 + \sigma^2},$$

90 where τ^2 is the between-patient variance and σ^2 is the within-patient variance. Since σ^2 defines the variance
 91 between individual measurements from a single patient, the ICC can be improved by using the mean of several
 92 measurements from a single patient rather than a single measurement. Doing so allows us to substitute σ^2
 93 with the variance of the sample mean, $\frac{\sigma^2}{n}$, giving us an ICC that is a function of the number of data points

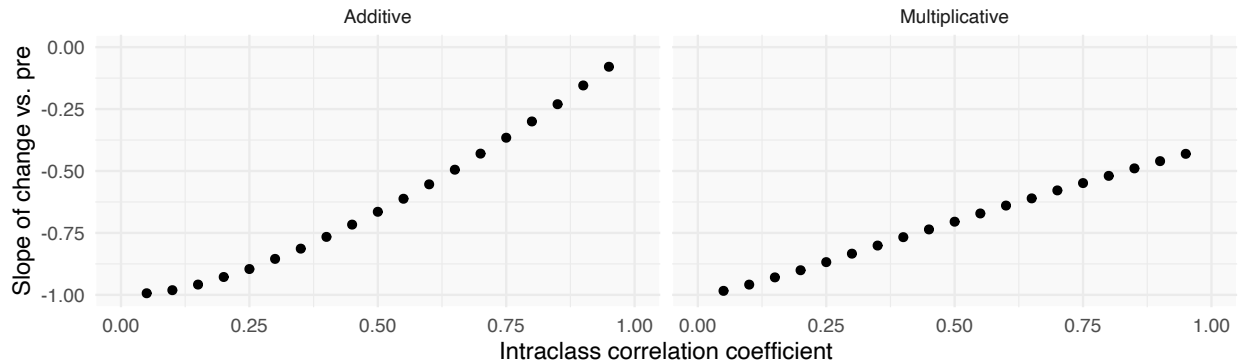


Figure 3: Simulations of additive and multiplicative changes reveal the effect of different intraclass correlation coefficients on the slope between change scores and pre-intervention scores. Additive effects have slopes that trend towards zero with increasing ICC’s, while multiplicative effects always have a negative slope no matter their ICC.

94 sampled from each patient,

$$\frac{\tau^2}{\tau^2 + \frac{\sigma^2}{n}}.$$

95 Note, this quantity approaches 1 (perfect reliability) as $n \rightarrow \infty$.

96 Importantly, the above concepts generalize to post-intervention scores as well. If we assume τ^2 and σ^2 do
 97 not change, and instead, there is a simple shift in mean scores without ceiling and floor effects, then the ICC
 98 also defines the Pearson correlation between pre- and post-intervention scores. The Pearson correlation is
 99 useful because it gives us direct insight into RTM—the slope between the pre-intervention scores and change
 100 scores approaches zero as the correlation between pre- and post-intervention scores approaches 1 (Figure 3).

101 This is depicted graphically in Figure 2b, which shows that those who have greater pre-intervention pain
 102 scores (x -axis) have smaller change scores (y -axis).

103 All of these properties come together and should be considered when statistically modeling pain relief
 104 and the effect of an intervention.

105 2.2 Multiplicative Model

106 The multiplicative model is still mathematically simple but its implications are more complex. If pain relief
 107 is multiplicative, then it can be modeled as a relative reduction; i.e., $\phi = \frac{\delta_i}{x_i}$. This would imply that each
 108 person’s post-intervention pain (y_i) is a fraction of their starting pain (x_i); i.e., $y_i = (\phi + 1)x_i$. However,
 109 ratios and relative reductions have unfavorable statistical properties. Instead, it is preferable to work on
 110 the log scale [Keene, 1995, Senn, 2006, Senn and Julious, 2009]. In particular, recall $\log \frac{y_i}{x_i} = \log y_i - \log x_i$,
 111 enabling us to linearize the multiplicative process. Similarly, from this, one may realize that it is natural to
 112 model multiplicative effects as being generated from *log*-normal distributions rather than normal distributions
 113 (Appendix B).

114 The implications of the log-normal distribution and its multiplicative properties are shown and described
 115 in Figure 2 and Figure 3. Note that the multiplicative pain reductions follow different distribution than
 116 additive effects owing to their errors compounding rather than adding. This results in a ‘fanning’ (or
 117 heteroscedasticity) of post-intervention scores as a function of greater pre-intervention scores Figure 2a.
 118 This is a hallmark of multiplicative processes that can be evaluated empirically. In addition to this fanning,
 119 it is quickly apparent that even with zero measurement error (Figure 3), multiplicative effects can look like
 120 RTM since greater pre-intervention scores will result in greater decreases in pain (Figure 2b). However,
 121 as opposed to additive processes in which greater pre-intervention scores are attributable to RTM (i.e.,
 122 measurement error), this relationship is indeed ‘real’ for multiplicative processes.

123 The multiplicative nature does not only apply to the relationship between pre- and post-intervention
 124 pain, but also the effect of a treatment. This is described in further detail in the next subsection.

2.3 Statistical Models of Pain Relief

Randomized controlled clinical trials aim to compare pain between two groups. To do so, investigators commonly compare the absolute or percent pain relief itself (e.g., a t -test on the change scores). However, such analyses are ill-conceived. Instead, especially for studies that record one or few follow-up measures (as opposed to time-series), it is recommended that the data-generating process be modeled using an analysis of covariance (ANCOVA) with pre-intervention scores as a covariate [Vickers and Altman, 2001, Senn, 2006]. The reasons for this are manifold:

1. The response variable in a statistical model should be the result of an experiment. Because patients enter studies with their baseline score, it is not the result of the experiment so it should not be treated as a dependent variable (e.g., like in a group \times time analysis of variance).
2. Accounting for RTM. Instead of a group \times time analysis of variance, one could perform a simple t -test on the change scores. However, such an analysis ignores RTM, and, especially in the case of baseline imbalances, can produce biased estimates. ANCOVA can adjust for such effects.
3. Improving statistical efficiency. ANCOVA has greater statistical efficiency, resulting in greater power and more precise intervals.
4. Post-intervention scores are arguably more interesting than change scores. Patients must live with the pain following the intervention, not the change in pain. However, regressing post-intervention pain *or* change in pain produces the same group effect [Senn, 2006].

These statistical and philosophical advantages are well-established in the biostatistics literature [Vickers and Altman, 2001, Vickers, 2001, Vickers, 2014, Frison and Pocock, 1992, Borm et al., 2007, Senn, 2006]. Note, the benefits of ANCOVA primarily apply to randomized studies, as ANCOVA may produce biased estimates in non-randomized studies depending on the allocation mechanism [Van Breukelen, 2006].

For the additive case, the ANCOVA model takes the form

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 g_i + \epsilon_i,$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ and g_i is dummy-coded for group (e.g., 0 = placebo and 1 = drug). β_2 is the effect of interest: the average difference in post-intervention pain scores between groups after adjusting for pre-intervention scores. β_1 will typically be < 1 , indicative of RTM, and the intercept may be nonsensical unless x_i is mean-centered. Of course, like any regression, one can add more covariates, especially those with prognostic value, which will further increase statistical efficiency.

The ANCOVA can also be generalized to the multiplicative case. Since multiplicative effects can be linearized by taking the log-transform, we can write the model as

$$y_i = B_0 \cdot x_i^{\beta_1} \cdot B_2^{g_i} \cdot E_i \tag{1}$$

$$= \exp \{ \beta_0 + \beta_1 \log x_i + \beta_2 g_i + \epsilon_i \} \tag{2}$$

$$\implies \log y_i = \beta_0 + \beta_1 \log x_i + \beta_2 g_i + \epsilon_i. \tag{3}$$

This model reveals a few things. First, in (1), residuals will compound with increasing values of the predicted y_i (i.e., \hat{y}_i). Indeed, this is consistent with what we observed in the simulations above, so this functional form can capture the compounding error. Second, in (3), both y_i and x_i are logged, so when $\beta_1 = 1$, it is equivalent to modeling the percent change; however, when $\beta_1 \neq 1$, there is a scaling to account for nonlinearities and RTM. Finally, B_2 is a multiplicative effect: when $B_2 = 1$, both groups are expected to have the same post-intervention score for a given pre-intervention score; when $B_2 > 1$, the experimental group is expected to have a greater post-intervention score for a given pre-intervention score; and so on. Since we are fitting β_2 rather than B_2 , the fit coefficient will be on the log scale, so exponentiating the coefficient will make it more interpretable despite the log scale having nicer mathematical properties. Note, even this multiplicative ANCOVA is more efficient than analyzing percent changes [Vickers, 2014].

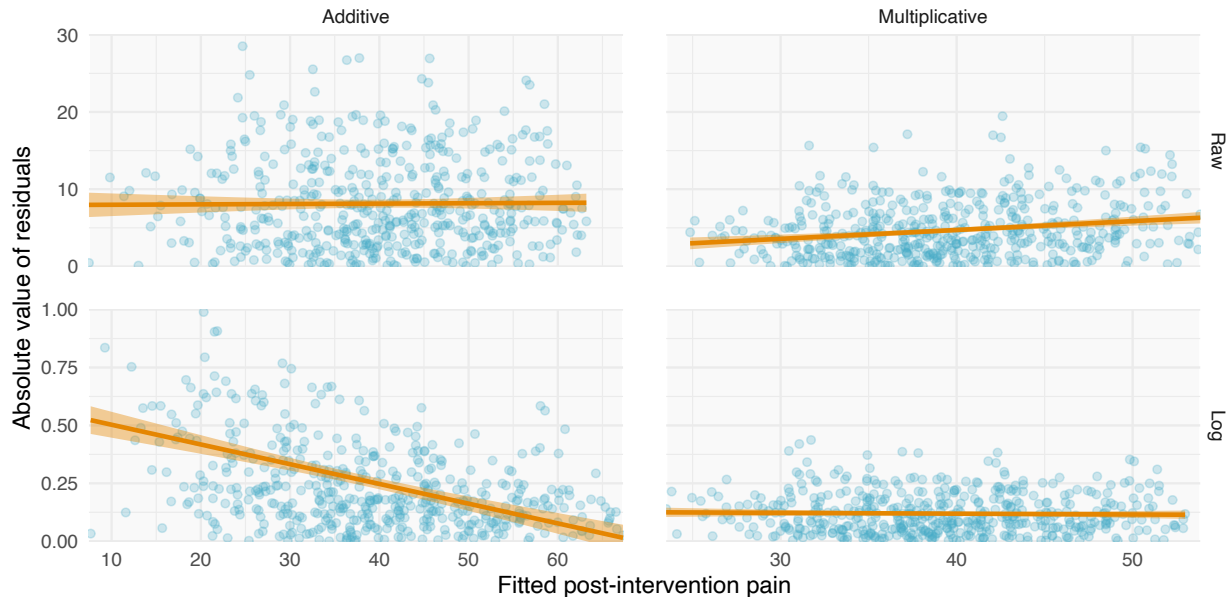


Figure 4: Simulations of additive and multiplicative changes reveal differential residual behavior for raw and log-transformed ANCOVA models. (Left) data generated with have an additive structure have homoscedastic residuals when fit with a standard ANCOVA (top) but heteroscedastic residuals when fit with a log-transformed ANCOVA (bottom). (Right) data generated with a multiplicative structure have homoscedastic residuals when fit on their raw scale (top) but homoscedastic residuals when log-transformed (bottom).

3 Empirical Data

As a proof of principle, we assessed the properties of four separate datasets. Two of the datasets were collected in patients with subacute back pain and the other two consist of patients with chronic back pain. Ideally, data are analyzed using intention-to-treat. However, here, we included individuals for whom we had enough ratings to complete our analyses as the data are being used for illustrative purposes and we are not looking to draw inferences.

3.1 Datasets

3.1.1 Placebo I (Chronic back pain)

Overview. The purpose of this study was to investigate factors associated with placebo analgesia in chronic pain patients [Vachon-Preseu et al., 2018]. This was the first trial designed to study chronic pain patients receiving placebo versus no treatment. The total duration of the study lasted ~ 15 months. Protocol and informed consent forms were approved by Northwestern University IRB and the study was conducted at Northwestern University (Chicago, IL, USA).

Participants. To meet inclusion criteria, individuals had to be 18 years or older with a history of lower back pain for at least 6 months. This pain should have been neuropathic (radiculopathy confirmed by physical examination was required), with no evidence of additional comorbid chronic pain, neurological, or psychiatric conditions. Individuals had to agree to stop any concomitant pain medications and had to be able to use a smartphone or computer to monitor pain twice a day. Additionally, the enrolled patients had to report a pain level of at least 5/10 during the screening interview, and their averaged pain level from the smartphone app needed to be higher than 4/10 during the baseline rating period before they were randomized into a treatment group. A total of 82 patients were randomized. Here, we include 18 participants from the no treatment group and 42 participants from the placebo group for whom we had complete rating data (cf. Supplementary Figure 1 in [Vachon-Preseu et al., 2018]).

188 *Pain data.* Data were collected using a custom pain rating phone app through which patients could rate
189 their pain (0–10 NRS). Patients were asked to enter their pain 2 times/day over the course of the entire
190 study. For the purposes of demonstration, here we averaged pain ratings within a single day.

191 3.1.2 Placebo II (Chronic back pain)

192 *Overview.* The purpose of this study was to validate a prognostic model for classifying chronic pain patients
193 based on their predicted improvement with placebo [Vachon-Preseu et al., 2021]. Protocol and informed
194 consent forms were approved by Northwestern University IRB and the study was conducted at Northwestern
195 University (Chicago, IL, USA).

196 *Participants.* Individuals with chronic low back pain were recruited for this study. Patients must have
197 had low back pain for at least 6 months, with or without symptoms of radiculopathy, a minimum VAS score
198 of 5/10 at the screening visit and a minimum average pain of 4/10 over a two-week period prior to their first
199 visit. A total of 94 patients were randomized to no treatment, placebo, or naproxen. Here, we include 12
200 participants from the no treatment group, 33 participants from the placebo group, and 35 participants from
201 the naproxen group for whom we had complete rating data (cf. Figure 1 in [Vachon-Preseu et al., 2021]).

202 *Pain data.* Data were collected using a custom pain rating phone app through which patients could rate
203 their pain (0–10 NRS), as in Placebo I. Patients were asked to enter their pain 2 times/day over the course
204 of the entire study. For the purposes of demonstration, here we averaged pain ratings within a single day.

205 3.1.3 Levodopa trial (Subacute back pain)

206 *Overview.* The purpose of this trial was to investigate whether levodopa (l-DOPA) can block patients’
207 transition from subacute to chronic back pain. This 24-week double-blind parallel randomized controlled
208 trial was conducted at Northwestern University (Chicago, IL, USA). Protocol and informed consent form
209 were approved by Northwestern University IRB as well as NIDCR/NIH. All enrolled participants provided
210 written informed consent. The trial was registered on ClinicalTrials.gov, under registry NCT01951105 and
211 is preprinted on medRxiv [Reckziegel et al., 2021].

212 *Participants.* Individuals with a recent onset of lower back pain were recruited. Criteria for enrollment
213 included history of lower back pain with duration between 4-20 weeks with signs and symptoms of radicu-
214 lopathy and average reported pain intensity > 4 (on an NRS scale from 0 to 10) on the week before baseline
215 assessments and the week preceding treatment start. Participants were randomized to one of three groups:
216 no treatment (completed $n=10$), naproxen + placebo ($n=28$), naproxen + l-DOPA/c-DOPA ($n=21$). Here,
217 we will use data from 47 patients who had complete rating data (naproxen + placebo = 27; naproxen +
218 l-DOPA/c-DOPA = 20) (cf. Figure 1b, [Reckziegel et al., 2021]).

219 *Pain data.* Data were collected using a custom pain rating phone app through which patients could rate
220 their pain (0–10 NRS). Patients were asked to enter their pain 3 times/day over the course of the entire
221 study (28 weeks). For the purposes of demonstration, here we averaged pain ratings within a single day.

222 3.1.4 Prospective cohort (Subacute back pain)

223 *Overview.* The purpose of this study was to identify predictive biomarkers to identify individuals who will
224 vs. will not recover from subacute back pain [Baliki et al., 2012]. Protocol and informed consent forms
225 were approved by Northwestern University IRB as well as NIDCR/NIH, and the study was conducted at
226 Northwestern University (Chicago, IL, USA). All enrolled participants provided written informed consent.
227 All participants were righthanded and were diagnosed by a clinician for back pain. An additional list of
228 criteria was imposed including: pain intensity $> 40/100$ on the visual analog scale (VAS) and duration < 16
229 weeks.

230 *Participants.* Eighty individuals with a recent onset (within 16 weeks) of lower back pain and an average
231 reported pain intensity $> 40/100$ (on the VAS) who completed at least three follow-up visits (i.e., 30 weeks
232 following the initial visit).

233 *Pain data.* Data were collected at 5 separate visits using the short form of the McGill Pain Questionnaire
234 (MPQ). The computed sensory and affective scores from the MPQ for each visit are used as individual pain
235 scores for each subject.

236 3.2 Data Properties

237 To evaluate whether each dataset was more compatible with an additive or multiplicative process, we con-
 238 ducted the same analyses from the Statistical Background section (Figures 2–4) on these data. In particular,
 239 we investigated properties of the raw and log-transformed data, in addition to the properties of ANCOVAs
 240 fit to the data. To do so, all data were converted to a 0–100 scale. Before log-transforming, we added 1 to
 241 the raw scores to avoid $\log(0) = \text{NaN}$. In doing so, we demonstrate how the aforementioned principles apply
 242 to real data.

243 All datasets have positive relationships between pre- and post-intervention scores (Figure 5, top row).
 244 Interestingly and in contrast to the other studies, the variance of the post-intervention scores in the levodopa
 245 trial appears to increase with greater pre-intervention scores, consistent with a multiplicative effect. Finally,
 246 with the exception of the prospective cohort study, there are negative relationships between changes in pain
 247 and pre-intervention scores. These negative relationships may be explained by multiplicative effects or RTM.
 248 Further examination is needed to ascertain the nature of these data.

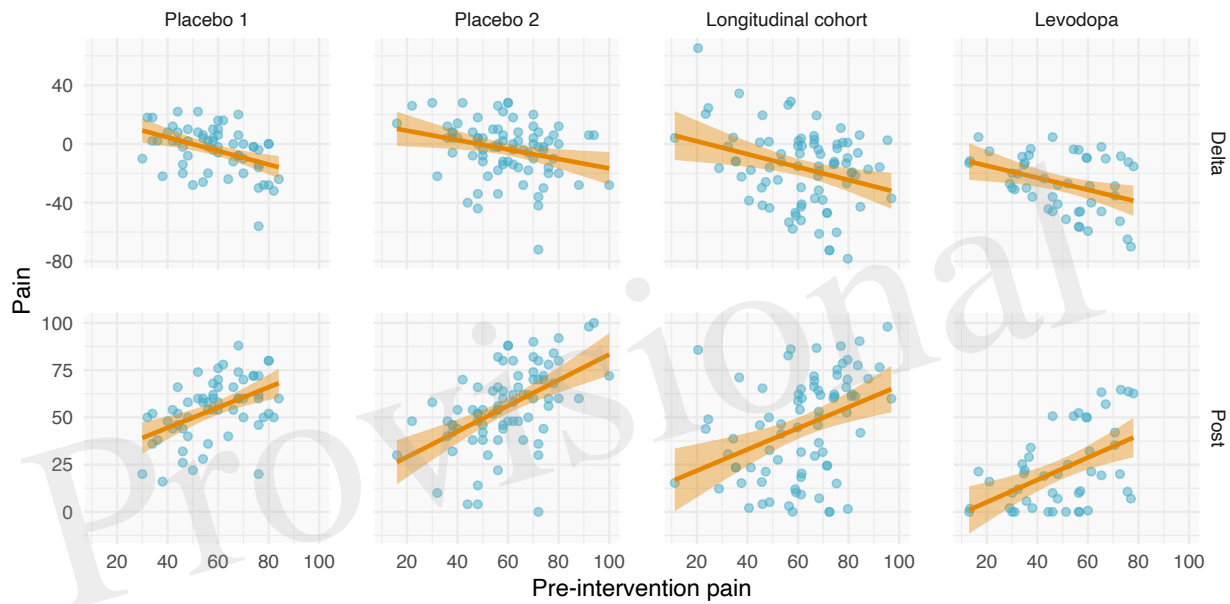


Figure 5: Relationships between pre-intervention scores and change scores (top) and pre-intervention scores (bottom). **(Top)** Relationship between pre-intervention scores and change scores. Note that most of the studies have a negative relationship. This could be explained by regression toward the mean *or* multiplicative effects, in addition to ceiling/floor effects. **(Bottom)** Relationship between pre-intervention and post-intervention pain scores across all studies. Each study shows a positive relationship between pre- and post-intervention scores; however, the Levodopa study appears to have greater variance in post-intervention scores with greater pre-intervention scores.

249 Including more points in the calculation of pre-intervention and post-intervention scores increases the
 250 ICC, thereby increasing the reliability and decreasing the effect of RTM (Figure 3). Since three of the four
 251 datasets contained ecological momentary assessments of pain, we were able to sample and average more
 252 than one point from the beginning and end of each study. We averaged an increasing number of a pre- and
 253 post-intervention points and recalculated the slope between change score and pre-intervention score (i.e.,
 254 plot from Figure 5, top). If the slopes strongly trend towards zero by increasing the number of points, this
 255 indicates that the data have additive properties. Slopes that stay negative regardless of increasing reliability
 256 (number of points) indicate that the data may be multiplicative. For the studies included in this analysis
 257 (Placebo I, Placebo II, Levodopa Trial), Placebo I and Placebo II’s slopes have slight upward trends: as the
 258 number of points in the calculation of pre-intervention and post-intervention scores increases, the negative
 259 slope due to RTM increases. In contrast, the Levodopa trial’s negative slopes remain stable (Figure 6). This

260 again hints at the notion that the levodopa trial’s data may be multiplicative, while Placebo I and Placebo
 261 II may be additive.

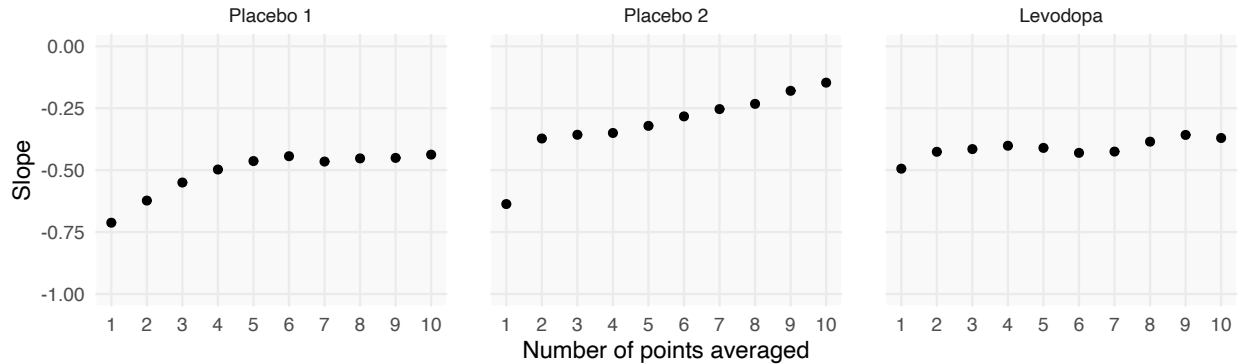


Figure 6: Increasing the number of points used for each patient’s pre- and post-intervention scores increases the slope between change scores and pre-intervention scores. Each patient’s pre- and post-intervention scores were calculated using the mean of x points. By averaging over more points, we should increase the intraclass correlation coefficient. Negative slopes between change scores and pre-intervention scores are indicative of one of two things: (1) regression toward the mean or (2) multiplicative effects. In the datasets that show evidence of being additive, we see marked increases in slopes, indicating that we are decreasing regression toward the mean by including more points. However, because the Levodopa Trial displays multiplicative properties, it is only minimally affected by adding more points.

262 Perhaps the most direct assessment of additive versus multiplicative properties is to model the data and
 263 assess the model fits. When assessing and utilizing a model, one should ensure that the model’s assumptions
 264 are met and that the model captures salient features of the data. Because multiplicative data-generating
 265 processes lead to compounding residuals, we can observe these effects when fitting ANCOVAs. In Figure 7,
 266 we focus specifically on the variance observed in Figure 5, illustrating the relationship between fitted values
 267 (using the ANCOVA models from Figure 5) and the absolute value of the residuals. As shown in Figure 2,
 268 multiplicative relationships possess higher variance as pre-intervention scores increase, compared to additive
 269 relationships which are homoscedastic. For this reason, we should observe a null correlation between fitted
 270 values and absolute residual error for data that have exhibited additive properties (Placebo I, Placebo II,
 271 Prospective Cohort) thus far, and observe a positive correlation between fitted values and absolute residual
 272 error for data that have exhibited multiplicative properties (Levodopa Trial). As predicted, the Placebo I,
 273 Placebo II, and Prospective Cohort data all display this additive quality, as their residual error does not
 274 increase as fitted values increase. In contrast, the Levodopa Trial data display multiplicative properties, as
 275 its residual error increases as fitted values increase. The description and analyses of these data can be seen
 276 below (Figure 7).

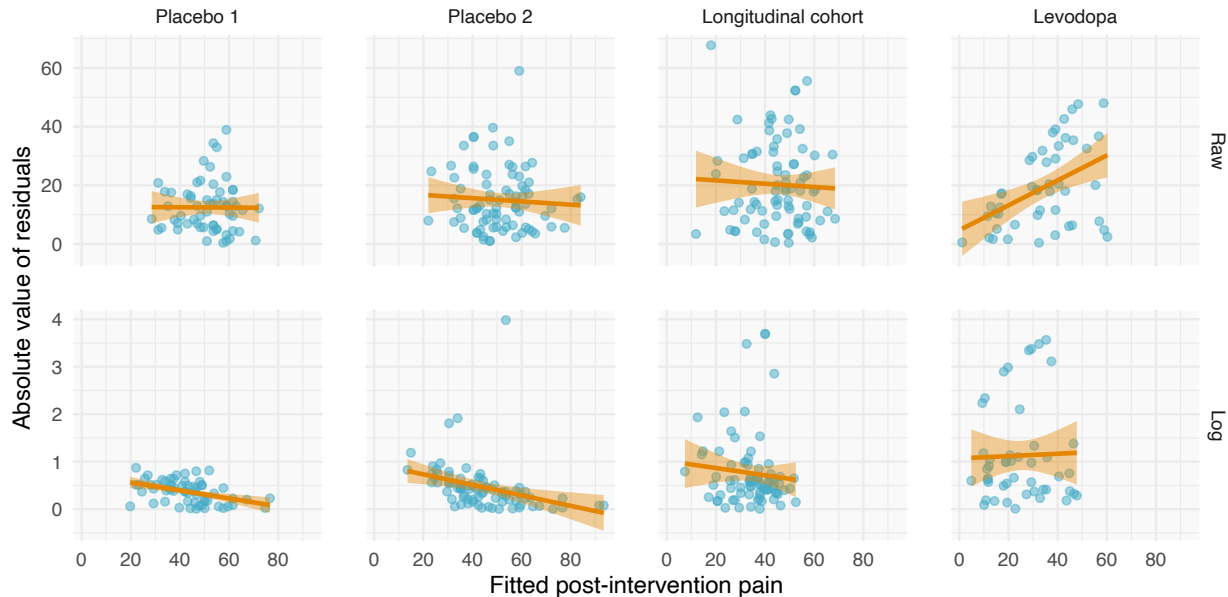


Figure 7: Absolute values of residuals from additive ANCOVA models. We fit an ANCOVA to each dataset using pre-intervention score and group membership as covariates. From these models, we plotted the absolute values of the residuals as a function of the fitted value. Additive models should be homoscedastic, meaning the magnitudes of the residuals do not change as a function of the response variable. However, multiplicative models have compounding error, such that if you fit them using an additive model, greater predicted values will be associated with larger magnitudes of residual error. Placebo I, Placebo II, and the Prospective Cohort study all exhibit features of additive data. However, the Levodopa Trial exhibits multiplicative properties, as evidenced by the increasing error residual magnitude with increasing fitted values.

277 From these plots, it is clear that the Placebo I, Placebo II, Prospective Cohort demonstrate additive
 278 properties while the Levodopa Trial demonstrates multiplicative properties. An understanding of these
 279 concepts and model assumptions have real implications. In Table 1, we include the average absolute (additive)
 280 and log-transformed (multiplicative) change in pain scores for each dataset. As an example, the effect of
 281 naproxen relative to no treatment in Placebo II is -15 (-27,-3) for the additive model but 0.7 (0.4, 1.1) for
 282 the multiplicative model. The 95% CI is much wider for the multiplicative model since it is misspecified,
 283 which in turn may lead an investigator or clinician to less certain conclusions about the treatment effect.

Dataset	Additive model (NRS), $\hat{\beta}$ (CI)	Multiplicative model (AU), $\hat{\beta}$ (CI)
Placebo I	-3 (-12, 5)	0.9 (0.8, 1.1)
Placebo II	Placebo: -9 (-21, 4) Naproxen: -15 (-27, -3)	Placebo: 0.8 (0.5, 1.3) Naproxen: 0.7 (0.4, 1.1)
Levodopa Trial	4 (-7, 15)	1.5 (0.7, 3.3)

Table 1: Additive and multiplicative effects by dataset. All effects were modeled using ANCOVA with pre-intervention scores as a covariate. Multiplicative effects use the log-transformed scores and represent the exponentiated coefficients which can be interpreted as the relative effect of treatment group versus the control group (e.g., post-intervention pain in the placebo group (Placebo I) will be 90% of the post-intervention pain in the no treatment group).

4 Discussion

Pain relief is a ubiquitous clinical trial outcome with direct treatment implications. Treatments that yield appreciable pain relief will be employed in the clinic, and findings from these trials may be communicated to patients. However, if data from trials are not properly modeled, then the resulting treatment effects may be both biased and highly variable, which in turn may mislead researchers, clinicians, and patients. In this theory-based paper, we have emphasized the difference between additive and multiplicative treating effects from mathematical, statistical, and empirical perspectives. It is clear that the assumptions behind these effects are not interchangeable and thus should be more thoughtfully considered when planning and analyzing clinical trial data. Moreover, how pain relief is conceptualized will propagate into the interpretation of effects, which we briefly discuss herein.

4.1 Minimal Clinically Important Differences

Pain intensity ratings can be difficult to interpret—they are a reductionist, unidimensional measurement intended to capture a single aspect of a private, complex, incommunicable experience [Chapman and Loeser, 1989, Turk and Melzack, 2011]. To help make sense of improvements, researchers and clinicians commonly rely on minimal clinically important differences (MCID). In clinical pain research, MCIDs are commonly derived by mapping changes in pain ratings onto a different scale, such as global impression of change [Farrar et al., 2001]. For example, what absolute change in NRS and relative change in NRS correspond to “much improved”? This mapping is then commonly used as a guidepost for interpreting other studies, and in some cases, individual patient changes [Smith et al., 2020].

Although commonly derived and used without justification, absolute and relative MCIDs are not interchangeable since they are mathematically incompatible with one another. Suppose patient A starts with an 8/10 pain and patient B starts with a 4/10 pain. If the treatment has an additive effect, both patients may improve by 2/10, but this would result in markedly different percent reductions: 25% and 50% for patients A and B, respectively. Farrar et al. [Farrar et al., 2001] suggest that an MCID for pain relief is 2/10 NRS or 30%; here, these would yield two different conclusions since both patients achieved a 2/10 decrease but only one patient achieved a 30% decrease. Much attention has been and continues to be given to both additive and multiplicative MCIDs without considering the conceptual difference between the two. This conceptual incompatibility needs to be reconciled if MCIDs are to be used in a meaningful way. However, there are also larger issues that warrant addressing.

Across studies and ignoring the numerical nature of treatment effects, MCIDs have a linear relationship with baseline pain ratings, with an x -intercept corresponding to roughly 30/100 and a slope of 1 (i.e., $\text{MCID} \approx \text{baseline} - 30$) [Olsen et al., 2018]. This relationship calls into question both absolute and relative MCIDs. If absolute MCIDs were valid, then we would expect the MCID to be constant across all baseline pain scores. If relative MCIDs were valid, then we would expect a y -intercept of 0 and a slope equal to the MCID. Rather, this relationship suggests MCIDs are more compatible with a post-intervention pain rather than change score, and this post-intervention pain is equal to 30/100. In other words, the MCID is the change in pain needed to obtain a 30/100. If true, this would be consistent with the idea that it is a patient’s pain, not change in pain, that is important.

More generally, MCIDs arguably represent a conflation of constructs. MCIDs typically involve dichotomizing a measurement by mapping it onto some other measurement using some loss function—a form of ‘dichotomania’ [Senn, 2005]. For example, researchers may threshold and dichotomize changes in VAS into improvement versus non-improvement using the global impression of change scale [Farrar et al., 2001]. This dichotomization of pain scores is then applied to other studies. Yet, such an approach is curious—it implies we are actually interested in global impression of change but use pain scores as a noisy proxy. If a researcher is interested in global impression of change, they should measure global impression of change as an outcome in their sample. Further, the ontological basis for dichotomous change scores is arguably ill-conceived. The insipid use of MCIDs in pain research and practice deserves greater scrutiny. From this perspective, it has been argued that greater context is needed in deriving metrics of clinical importance [Ferreira et al., 2012, Ferreira, 2018] for which decision theory may provide a rigorous foundation.

In addition to using MCIDs for interpreting findings, researchers have used MCIDs for ‘responder analysis’. For example, a researcher may split patients into groups of ‘responders’ and ‘non-responders’ based

335 on whether their change in pain exceed the MCID (see Section 4.5 in [Smith et al., 2020]). However, such
336 analyses have undesirable properties on both the individual and group levels. On the individual level, in-
337 ferences cannot be made regarding response magnitude for several reasons. First, individual counterfactuals
338 are not observed in parallel group trials; for example, we do not know what what an individual’s pain would
339 have been had they been randomized to the placebo group instead of the drug group. An individual’s
340 observed improvement or worsening may have been due to the intervention or alternatively, RTM, natural
341 history, or some other unmeasured, stochastic process. Second, the individual may not reliably attain the
342 same improvement each time the trial is performed; for example, 60% of individuals may respond 100%
343 of the time *or* 100% of individuals may respond 60% of the time (or some mixture of the two). Third,
344 this dichotomization assumes an improvement of, say, 30% and 100% are equivalent, and similarly, that
345 an improvement of 29% and 0% are equivalent (assuming MCID = 30%) by treating improvements as a
346 binary step function rather than continuous—such an assumption strains credulity. These issues have been
347 previously discussed in great detail [Senn, 2001, Senn, 2004, Senn, 2016, Senn, 2018]. On the group level,
348 dichotomizing individual responses turns each patient’s pain improvement into a 0 (‘non-responder’) or 1
349 (‘responder’), which discards information and, in turn, markedly decreases statistical efficiency and power
350 [Cohen, 1983, Altman and Royston, 2006]. Thus, the dichotomization of improvements is arguably unethical
351 since it discards information, effectively decreasing the sample size [Cohen, 1983] and, in turn, the ability
352 to quantify (or rule out) meaningful intervention effects. Rather than being treated as an analytical tool,
353 MCIDs are perhaps better viewed from an interpretive and decision-making perspective.

354 Notwithstanding MCID’s limitations, it is perhaps most useful at the planning stage of clinical research.
355 A clinically important difference is just one approach to justifying an effect size of interest for a study
356 [Cook et al., 2018], which may be used for sample size calculations or stopping rules in adaptive trials.
357 However, beyond planning, dichotomizing trial and especially individual patient outcomes using an MCID
358 is a questionable practice that commonly ignores context and variability [Senn and Julious, 2009].

359 4.2 Scale Assumptions

360 Psychological measurement scales have a rich history across the fields of psychometrics and psychophysics
361 [Stevens, 2017]. Anchors determine the extremes within which a participant must rate their experience,
362 ultimately constraining the measurement construct and how accurately participants understand what they
363 are rating [Yokobe et al., 2014]. Bounded by these anchors, the measurements themselves can be on one
364 of a number of scales: nominal, ordinal, interval, ratio, and absolute. Nominal scales assume a one-
365 to-one mapping between the desired quantity x' and the measured quantity x ; ordinal scales assume a
366 monotonic mapping; interval scales assume an affine mapping ($x' = ax + b$); ratio scales assume a lin-
367 ear mapping with an absolute zero ($x' = ax$); and absolute scales assume a perfect mapping ($x' = x$)
368 [Bolanowski Jr and Gescheider, 2013]. Several renowned psychophysicists have argued—not without criti-
369 cism [Ellermeier and Faulhammer, 2000, Zimmer, 2005]—that perceptual ratings are or can easily be con-
370 verted to ratio scale [Stevens, 2017, Bolanowski Jr and Gescheider, 2013]. Importantly, the additive and
371 multiplicative models rely on interval and ratio assumptions, respectively. Thus, the validity of these as-
372 sumptions for clinical pain must be considered.

373 The numerical nature of clinical pain is an open, controversial, and perhaps unanswerable question. Early
374 psychophysics work argues that VAS and NRS pain scales are ratio for both experimental and clinical pain. In
375 1983, Price et al. [Price et al., 1983] used cross-modality matching to argue that clinical pain, like heat pain,
376 is a ratio scale. However, by mapping clinical pain onto heat pain, this finding is arguably tautological—
377 they assessed whether clinical pain-matched heat pain follows the same power law as heat pain. Others
378 have used item-response theory to argue that pain ratings are ordinal scale (nonlinear) rather than ratio or
379 interval scale [Kersten et al., 2014]. Since the authors used unidimensional measures and a Rasch model, this
380 conclusion is based on stationarity assumptions and ratings’ reliability, which are not necessary conditions
381 for interval or ratio scales. Although the perceptual ratings from psychophysics are undoubtedly related to
382 clinical pain, assessing the measurement properties of clinical pain is much more complex since we cannot
383 precisely control the sensory input. Thus, clinical pain measurement scale assumptions arguably cannot be
384 rigorously evaluated, reinforcing that they are indeed assumptions. However, the strength of assumption
385 varies, with interval scales (additive) having weaker assumptions than ratio scales (multiplicative). The
386 assumptions a researcher makes directly affects the model they should choose.

4.3 Statistical Modeling and Applications

The choice of a statistical model can greatly affect the inferences drawn from the same dataset. Here, we observed that applying a multiplicative model to a dataset that exhibits additive properties can create wide CIs, making it difficult to interpret the results of an experiment (Table 1). This is consistent with the idea that a properly specified model will be more statistically efficient [Vickers, 2014], and perhaps most importantly, they better represent the underlying data.

We presented two ways of modeling data: additively and multiplicatively. Both rely on ANCOVA, with the former using raw pain scores and the latter using log-transformed pain scores. These models have different assumptions about the underlying data and, as a result, have different interpretations. If authors feel the linearity and ratio assumptions are too strict, there are other models that can be used; e.g., ordinal regression and semiparametric (or nonparametric) ANCOVA [Harrell, 2017], in addition to intensive longitudinal and time-series analysis [Fitzmaurice, 2011]. Indeed, there are good examples in the pain literature of ANCOVA-type models being implemented with more complicated data structures (e.g., multiple study endpoints, see [Mathieson et al., 2017]). In any case, researchers should be aware of the assumptions of their statistical models of the properties of their data, and of course, researchers are encouraged to collaborate with statisticians [Sainani et al., 2021].

4.4 Recommendations

We have clearly demonstrated the mathematical, conceptual, and interpretive differences between additive and multiplicative effects. From this explication, there are tangible takeaways and recommendations for clinical researchers. Specifically, we suggest that researchers include and consider the following:

1. When deciding which metric to use—absolute pain decreases or percent pain decreases—use the data as a guide unless there is a principled reason to choose one or the other. Since it is unclear what influences the presence of additive or multiplicative characteristics in pain data, it is safer to use the metric that exhibits the properties of the data accurately. Table 2 summarizes the differences between additive and multiplicative properties. In time, we may develop a better understanding of pain conditions and improvements such that more general recommendations can be provided. We view this as being no different than checking statistical model assumptions.

Table 2: Hallmarks of additive and multiplicative effects.

Plot	Additive	Multiplicative
Slope of change score vs. pre-intervention score (y) vs. number of points (x)	Slopes approach zero as the number of points utilized in calculating pre- and post-intervention pain scores increases by increasing ICC (Figure 3, left).	Slopes increase minimally with increasing number of points (Figure 3, right).
Absolute value of residuals (y) vs. fitted values (x)	No relationship between absolute residual error and fitted (post-intervention) values.	Positive, heteroscedastic relationship between absolute residual error and fitted (post-intervention) values.

2. When reporting descriptive statistics, use the arithmetic mean to calculate between-subject (average) intervention for additive data; conversely, use geometric mean for multiplicative data.
3. Ensure that patients' pre-intervention scores are heterogeneous for drawing conclusions about the nature of the data. By including a wide range of pre-intervention scores, it makes the additive or multiplicative properties more apparent. If the data are not heterogeneous, false conclusions may be made about the data's additive or multiplicative properties.

5 Conclusion

The properties of changes in self-reported pain are commonly implicitly assumed to be additive, multiplicative, or are conflated. Ignoring the properties of pain relief can result in model misspecification, in turn leading to bias and statistical inefficiency. These errors further propagate into metrics such as minimal clinically important differences. We contend that more attention should be paid to the statistical properties of pain relief to ensure model assumptions are met. By paying closer attention to these properties, we can gain more insight from and make better use of data from pain clinical trials.

A Data Generating Processes

A.1 Additive Model

The additive model can be conceptualized hierarchically. First, we will assume each individual's average pre-intervention pain, α_i for patient i , is sampled from a larger population,

$$\alpha_i \sim \mathcal{N}(\mu, \tau^2).$$

Since α_i represents an individual's *average* pre-intervention pain, it is a latent construct and ignores measurement error and natural pain variability; for example, minute-to-minute, hour-to-hour, and day-to-day fluctuations in pain intensity. In actuality, an experiment will sample an individual's pain ratings and will be affected by measurement error. Thus, a given measurement of a patient's pre-intervention pain will be

$$x_{ij} = \alpha_i + \epsilon_{ij},$$

where $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ for measurement j from patient i , assuming all patients have the same within-patient variability (Figure 1). If we sample and average n measurements from patient i , we obtain

$$x_{i.} \sim \mathcal{N}\left(\alpha_i, \frac{\sigma^2}{n}\right).$$

Similarly, assuming homogeneous improvement and treatment effects, the average post-intervention pain rating for patient i is

$$y_{i.} \sim \mathcal{N}\left(\alpha_i + \delta + \theta g_i, \frac{\sigma^2}{n}\right),$$

where δ is the improvement in the control group, θ the treatment effect of interest, and g_i is a dummy variable for group (0 = control; 1 = intervention). Without loss of generality via the additive assumption of treatment effects, we will ignore treatment groups (θ) to simplify the problem and describe the properties of these distributions, giving us the simplified post-intervention pain distribution

$$y_{i.} \sim \mathcal{N}\left(\alpha_i + \delta, \frac{\sigma^2}{n}\right).$$

For both the pre and post model, the intraclass correlation coefficient (ICC) is

$$ICC = \frac{\tau^2}{\tau^2 + \frac{\sigma^2}{n}},$$

which is also the correlation between pre- and post-intervention scores. Luckily, ICC is sensitive to the number of data points from which each patient's pre- and post-intervention mean pain scores are calculated,

$$\lim_{n \rightarrow \infty} \frac{\sigma^2}{n} = 0 \implies \lim_{n \rightarrow \infty} \frac{\tau^2}{\tau^2 + \frac{\sigma^2}{n}} = 1.$$

With more data points, the slope attributable to RTM disappears. Since the ICC is equivalent to a Pearson's r in this case, we can write the joint pre-post distribution of averaged pain scores can be written as a multivariate normal,

$$\begin{pmatrix} x_{i.} \\ y_{i.} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ \mu + \delta \end{pmatrix}, \begin{bmatrix} \tau^2 + \frac{\sigma^2}{n} & \tau^2 \\ \tau^2 & \tau^2 + \frac{\sigma^2}{n} \end{bmatrix}\right).$$

449 A.2 Multiplicative Model

450 The log-normal distribution is an exponentiated normal distribution, meaning the log of the log-normal
451 distribution is a normal distribution. Therefore, we have

$$\log \alpha_i \sim \mathcal{N} \left(\log \left(\frac{\mu^2}{\sqrt{\mu^2 + \tau^2}} \right), \log \left(1 + \frac{\tau^2}{\mu^2} \right) \right).$$

452 And like the additive case, a single pre-intervention score j pain for patient i can be described as being
453 centered around their individual mean,

$$\log x_{ij} \sim \mathcal{N} \left(\log \alpha_i, \frac{\sigma}{\mu} \right).$$

454 Similarly, a patient's post-intervention pain is scaled rather than shifted by the change in pain, δ ,

$$\log y_{ij} \sim \mathcal{N} \left(\log \alpha_i + \log \left(1 + \frac{\delta}{\mu} \right), \frac{\sigma}{\mu} \right).$$

455 B Acknowledgments

456 This work is funded by the National Institutes of Health (1P50DA044121-01A1). This material is based
457 upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No.
458 DGE-1324585.

459 References

- 460 [Altman and Royston, 2006] Altman, D. G. and Royston, P. (2006). The cost of dichotomising continuous
461 variables. *BMJ*, 332(7549):1080.
- 462 [Baliki et al., 2012] Baliki, M. N., Petre, B., Torbey, S., Herrmann, K. M., Huang, L., Schnitzer, T. J.,
463 Fields, H. L., and Apkarian, A. V. (2012). Corticostriatal functional connectivity predicts transition to
464 chronic back pain. *Nat Neurosci*, 15(8):1117–1119.
- 465 [Bolanowski Jr and Gescheider, 2013] Bolanowski Jr, S. J. and Gescheider, G. A. (2013). *Ratio scaling of*
466 *psychological magnitude: In honor of the memory of SS Stevens*. Psychology Press.
- 467 [Borm et al., 2007] Borm, G. F., Fransen, J., and Lemmens, W. A. (2007). A simple sample size formula for
468 analysis of covariance in randomized clinical trials. *Journal of clinical epidemiology*, 60(12):1234–1238.
- 469 [Chapman and Loeser, 1989] Chapman, C. and Loeser, J. (1989). *Issues in pain measurement*. Raven Press,
470 New York.
- 471 [Cohen, 1983] Cohen, J. (1983). The cost of dichotomization. *Applied psychological measurement*, 7(3):249–
472 253.
- 473 [Cook et al., 2018] Cook, J. A., Julious, S. A., Sones, W., Hampson, L. V., Hewitt, C., Berlin, J. A., Ashby,
474 D., Emsley, R., Fergusson, D. A., Walters, S. J., Wilson, E. C. F., MacLennan, G., Stallard, N., Rothwell,
475 J. C., Bland, M., Brown, L., Ramsay, C. R., Cook, A., Armstrong, D., Altman, D., and Vale, L. D.
476 (2018). DELTA2 guidance on choosing the target difference and undertaking and reporting the sample
477 size calculation for a randomised controlled trial. *BMJ*, 363:k3750.
- 478 [Ellermeier and Faulhammer, 2000] Ellermeier, W. and Faulhammer, G. (2000). Empirical evaluation of
479 axioms fundamental to Stevens's ratio-scaling approach: I. Loudness production. *Percept Psychophys*,
480 62(8):1505–1511.

- 481 [Farrar et al., 2001] Farrar, J. T., Young, J. P., LaMoreaux, L., Werth, J. L., and Poole, M. R. (2001).
482 Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating
483 scale. *Pain*, 94(2):149–158.
- 484 [Ferreira, 2018] Ferreira, M. (2018). Research Note: The smallest worthwhile effect of a health intervention.
485 *J Physiother*, 64(4):272–274.
- 486 [Ferreira et al., 2012] Ferreira, M. L., Herbert, R. D., Ferreira, P. H., Latimer, J., Ostelo, R. W., Nascimento,
487 D. P., and Smeets, R. J. (2012). A critical review of methods used to determine the smallest worthwhile
488 effect of interventions for low back pain. *J Clin Epidemiol*, 65(3):253–261.
- 489 [Fitzmaurice, 2011] Fitzmaurice, G. (2011). *Applied longitudinal analysis*. Wiley, Hoboken, N.J.
- 490 [Frison and Pocock, 1992] Frison, L. and Pocock, S. J. (1992). Repeated measures in clinical trials: analysis
491 using mean summary statistics and its implications for design. *Statistics in medicine*, 11(13):1685–1704.
- 492 [Gaskin and Richard, 2012] Gaskin, D. J. and Richard, P. (2012). The economic costs of pain in the United
493 States. *J Pain*, 13(8):715–724.
- 494 [Harrell, 2017] Harrell, F. E. (2017). Regression modeling strategies. *Bios*, 330(2018):14.
- 495 [Keene, 1995] Keene, O. N. (1995). The log transformation is special. *Statistics in medicine*, 14(8):811–819.
- 496 [Kersten et al., 2014] Kersten, P., White, P. J., and Tennant, A. (2014). Is the pain visual analogue scale
497 linear and responsive to change? An exploration using Rasch analysis. *PLoS One*, 9(6):e99485.
- 498 [Mathieson et al., 2017] Mathieson, S., Maher, C. G., McLachlan, A. J., Latimer, J., Koes, B. W., Hancock,
499 M. J., Harris, I., Day, R. O., Billot, L., Pik, J., Jan, S., and Lin, C. C. (2017). Trial of Pregabalin for
500 Acute and Chronic Sciatica. *N Engl J Med*, 376(12):1111–1120.
- 501 [Mntyselk et al., 2001] Mntyselk, P., Kumpusalo, E., Ahonen, R., Kumpusalo, A., Kauhanen, J., Viinamki,
502 H., Halonen, P., and Takala, J. (2001). Pain as a reason to visit the doctor: a study in Finnish primary
503 health care. *Pain*, 89(2-3):175–180.
- 504 [Nasution et al., 2018] Nasution, I. K., Lubis, N. D. A., Amelia, S., and Hocin, K. (2018). The correlation of
505 pain intensity and quality of life in chronic LBP patients in adam malik general hospital. *IOP Conference*
506 *Series: Earth and Environmental Science*, 125:012183.
- 507 [Olsen et al., 2018] Olsen, M. F., Bjerre, E., Hansen, M. D., Tendal, B., Hilden, J., and Hrbjartsson, A.
508 (2018). Minimum clinically important differences in chronic pain vary considerably by baseline pain and
509 methodological factors: systematic review of empirical studies. *J Clin Epidemiol*, 101:87–106.
- 510 [Price et al., 1983] Price, D. D., McGrath, P. A., Rafii, A., and Buckingham, B. (1983). The validation of
511 visual analogue scales as ratio scale measures for chronic and experimental pain. *Pain*, 17(1):45–56.
- 512 [Reckziegel et al., 2021] Reckziegel, D., Ttreault, P., Ghantous, M., Wakaizumi, K., Petre, B., Huang, L.,
513 Jabakhanji, R., Abdullah, T., Vachon-Presseau, E., Berger, S., Baria, A., Griffith, J. W., Baliki, M. N.,
514 Schnitzer, T. J., and Apkarian, A. V. (2021). Sex-Specific Pharmacotherapy for Back Pain: A Proof-of-
515 Concept Randomized Trial. *Pain Ther*.
- 516 [Sainani et al., 2021] Sainani, K. L., Borg, D. N., Caldwell, A. R., Butson, M. L., Tenan, M. S., Vickers,
517 A. J., Vigotsky, A. D., Warmenhoven, J., Nguyen, R., Lohse, K. R., Knight, E. J., and Bargary, N.
518 (2021). Call to increase statistical collaboration in sports science, sport and exercise medicine and sports
519 physiotherapy. *Br J Sports Med*, 55(2):118–122.
- 520 [Senn, 2001] Senn, S. (2001). Individual therapy: new dawn or false dawn? *Drug Information Journal*,
521 35(4):1479–1494.
- 522 [Senn, 2004] Senn, S. (2004). Individual response to treatment: is it a valid assumption? *BMJ*,
523 329(7472):966–968.

- 524 [Senn, 2006] Senn, S. (2006). Change from baseline and analysis of covariance revisited. *Stat Med*,
525 25(24):4334–4344.
- 526 [Senn, 2016] Senn, S. (2016). Mastering variation: variance components and personalised medicine. *Stat*
527 *Med*, 35(7):966–977.
- 528 [Senn, 2018] Senn, S. (2018). Statistical pitfalls of personalized medicine. *Nature*, 563(7733):619–621.
- 529 [Senn and Julious, 2009] Senn, S. and Julious, S. (2009). Measurement in clinical trials: a neglected issue
530 for statisticians? *Statistics in medicine*, 28(26):3189–3209.
- 531 [Senn, 2005] Senn, S. J. (2005). Dichomania: An obsessive compulsive disorder that is badly affecting the
532 quality of analysis of pharmaceutical trials. In *Proceedings of the International Statistical Institute, 55th*
533 *Session*, Sydney.
- 534 [Smith et al., 2020] Smith, S. M., Dworkin, R. H., Turk, D. C., McDermott, M. P., Eccleston, C., Farrar,
535 J. T., Rowbotham, M. C., Bhagwagar, Z., Burke, L. B., Cowan, P., et al. (2020). Interpretation of chronic
536 pain clinical trial outcomes: Impact recommended considerations. *Pain*, 161(11):2446–2461.
- 537 [St Sauver et al., 2013] St Sauver, J. L., Warner, D. O., Yawn, B. P., Jacobson, D. J., McGree, M. E.,
538 Pankratz, J. J., Melton, L. J., Roger, V. L., Ebbert, J. O., and Rocca, W. A. (2013). Why patients visit
539 their doctors: assessing the most prevalent conditions in a defined American population. *Mayo Clin Proc*,
540 88(1):56–67.
- 541 [Stevens, 2017] Stevens, S. S. (2017). *Psychophysics: Introduction to its perceptual, neural and social*
542 *prospects*. Routledge.
- 543 [Turk and Melzack, 2011] Turk, D. and Melzack, R. (2011). *Handbook of pain assessment*. Guilford Press,
544 New York.
- 545 [Vachon-Preseu et al., 2021] Vachon-Preseu, E., Abdullah, T. B., Berger, S. E., Huang, L., Griffith,
546 J. W., Schnitzer, T. J., and Apkarian, A. V. (2021). Validating a biosignature predicting placebo pill
547 response in chronic pain in the settings of a randomized controlled trial. *Pain*.
- 548 [Vachon-Preseu et al., 2018] Vachon-Preseu, E., Berger, S. E., Abdullah, T. B., Huang, L., Cecchi, G. A.,
549 Griffith, J. W., Schnitzer, T. J., and Apkarian, A. V. (2018). Brain and psychological determinants of
550 placebo pill response in chronic pain patients. *Nat Commun*, 9(1):3397.
- 551 [Van Breukelen, 2006] Van Breukelen, G. J. (2006). ANCOVA versus change from baseline: more power in
552 randomized studies, more bias in nonrandomized studies. *J Clin Epidemiol*, 59(9):920–925.
- 553 [Vickers, 2001] Vickers, A. J. (2001). The use of percentage change from baseline as an outcome in a
554 controlled trial is statistically inefficient: a simulation study. *BMC medical research methodology*, 1(1):1–
555 4.
- 556 [Vickers, 2014] Vickers, A. J. (2014). Change/percent change from baseline. *Wiley StatsRef: Statistics*
557 *Reference Online*.
- 558 [Vickers and Altman, 2001] Vickers, A. J. and Altman, D. G. (2001). Statistics notes: Analysing controlled
559 trials with baseline and follow up measurements. *BMJ*, 323(7321):1123–1124.
- 560 [Vos et al., 2020] Vos, T., Lim, S. S., Abbafati, C., Abbas, K. M., Abbasi, M., Abbasifard, M., Abbasi-
561 Kangevari, M., Abbastabar, H., Abd-Allah, F., Abdelalim, A., Abdollahi, M., Abdollahpour, I., Abolhas-
562 sani, H., Aboyans, V., Abrams, E. M., Abreu, L. G., Abrigo, M. R. M., Abu-Raddad, L. J., Abushouk,
563 A. I., Acebedo, A., Ackerman, I. N., Adabi, M., Adamu, A. A., Adebayo, O. M., Adekanmbi, V., Adelson,
564 J. D., Adetokunboh, O. O., Adham, D., Afshari, M., Afshin, A., Agardh, E. E., Agarwal, G., Agesa,
565 K. M., Aghaali, M., Aghamir, S. M. K., Agrawal, A., Ahmad, T., Ahmadi, A., Ahmadi, M., Ahmadi, H.,
566 Ahmadpour, E., Akalu, T. Y., Akinyemi, R. O., Akinyemiju, T., Akombi, B., Al-Aly, Z., Alam, K.,
567 Alam, N., Alam, S., Alam, T., Alanzi, T. M., Albertson, S. B., Alcalde-Rabanal, J. E., Alema, N. M., Ali,

568 M., Ali, S., Alicandro, G., Alijanzadeh, M., Alinia, C., Alipour, V., Aljunid, S. M., Alla, F., Allebeck, P.,
569 Almasi-Hashiani, A., Alonso, J., Al-Raddadi, R. M., Altirkawi, K. A., Alvis-Guzman, N., Alvis-Zakzuk,
570 N. J., Amini, S., Amini-Rarani, M., Aminorroaya, A., Amiri, F., Amit, A. M. L., Amugsi, D. A., Amul, G.
571 G. H., Anderlini, D., Andrei, C. L., Andrei, T., Anjomshoa, M., Ansari, F., Ansari, I., Ansari-Moghaddam,
572 A., Antonio, C. A. T., Antony, C. M., Antriyandarti, E., Anvari, D., Anwer, R., Arabloo, J., Arab-Zozani,
573 M., Aravkin, A. Y., Ariani, F., rnlv, J., Aryal, K. K., Arzani, A., Asadi-Aliabadi, M., Asadi-Pooya, A. A.,
574 Asghari, B., Ashbaugh, C., Atnafu, D. D., Atre, S. R., Ausloos, F., Ausloos, M., Ayala Quintanilla, B. P.,
575 Ayano, G., Ayanore, M. A., Aynalem, Y. A., Azari, S., Azarian, G., Azene, Z. N., Babae, E., Badawi, A.,
576 Bagherzadeh, M., Bakhshaei, M. H., Bakhtiari, A., Balakrishnan, S., Balalla, S., Balassyano, S., Banach,
577 M., Banik, P. C., Bannick, M. S., Bante, A. B., Baraki, A. G., Barboza, M. A., Barker-Collo, S. L.,
578 Barthelemy, C. M., Barua, L., Barzegar, A., Basu, S., Baune, B. T., Bayati, M., Bazmandegan, G., Bedi,
579 N., Beghi, E., Bjot, Y., Bello, A. K., Bender, R. G (2020). Global burden of 369 diseases and injuries
580 in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study
581 2019. *Lancet*, 396(10258):1204–1222.

582 [Yazdi-Ravandi et al., 2013] Yazdi-Ravandi, S., Taslimi, Z., Jamshidian, N., Saberi, H., Shams, J., and
583 Haghparast, A. (2013). Prediction of Quality of life by Self-Efficacy, Pain Intensity and Pain Duration in
584 Patient with Pain Disorders. *Basic Clin Neurosci*, 4(2):117–124.

585 [Yokobe et al., 2014] Yokobe, J., Kitahara, M., Matsushima, M., and Uezono, S. (2014). Preference for
586 different anchor descriptors on visual analogue scales among Japanese patients with chronic pain. *PLoS*
587 *One*, 9(6):e99891.

588 [Zimmer, 2005] Zimmer, K. (2005). Examining the validity of numerical ratios in loudness fractionation.
589 *Percept Psychophys*, 67(4):569–579.

Figure 01.JPEG

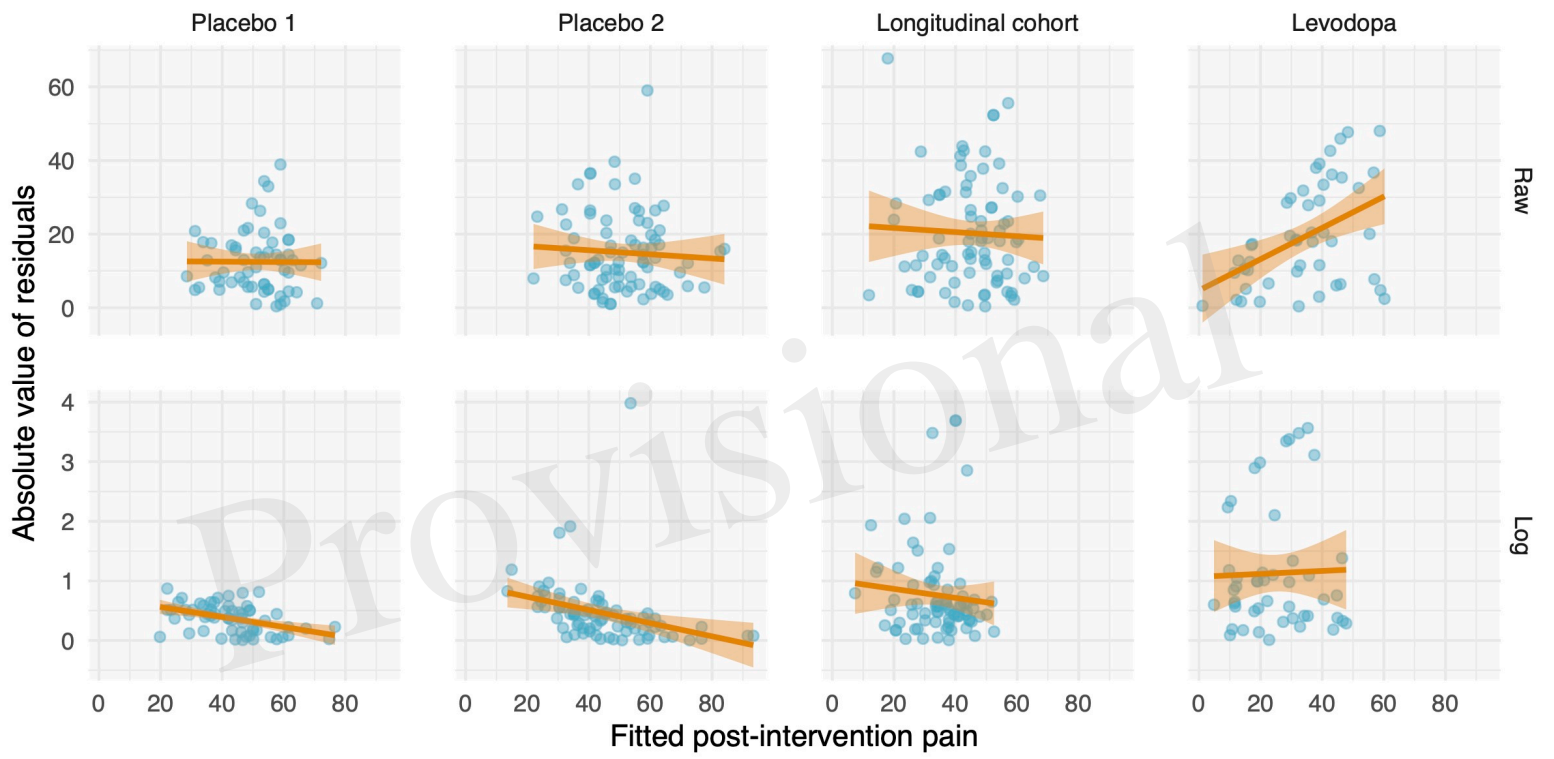


Figure 02.JPEG

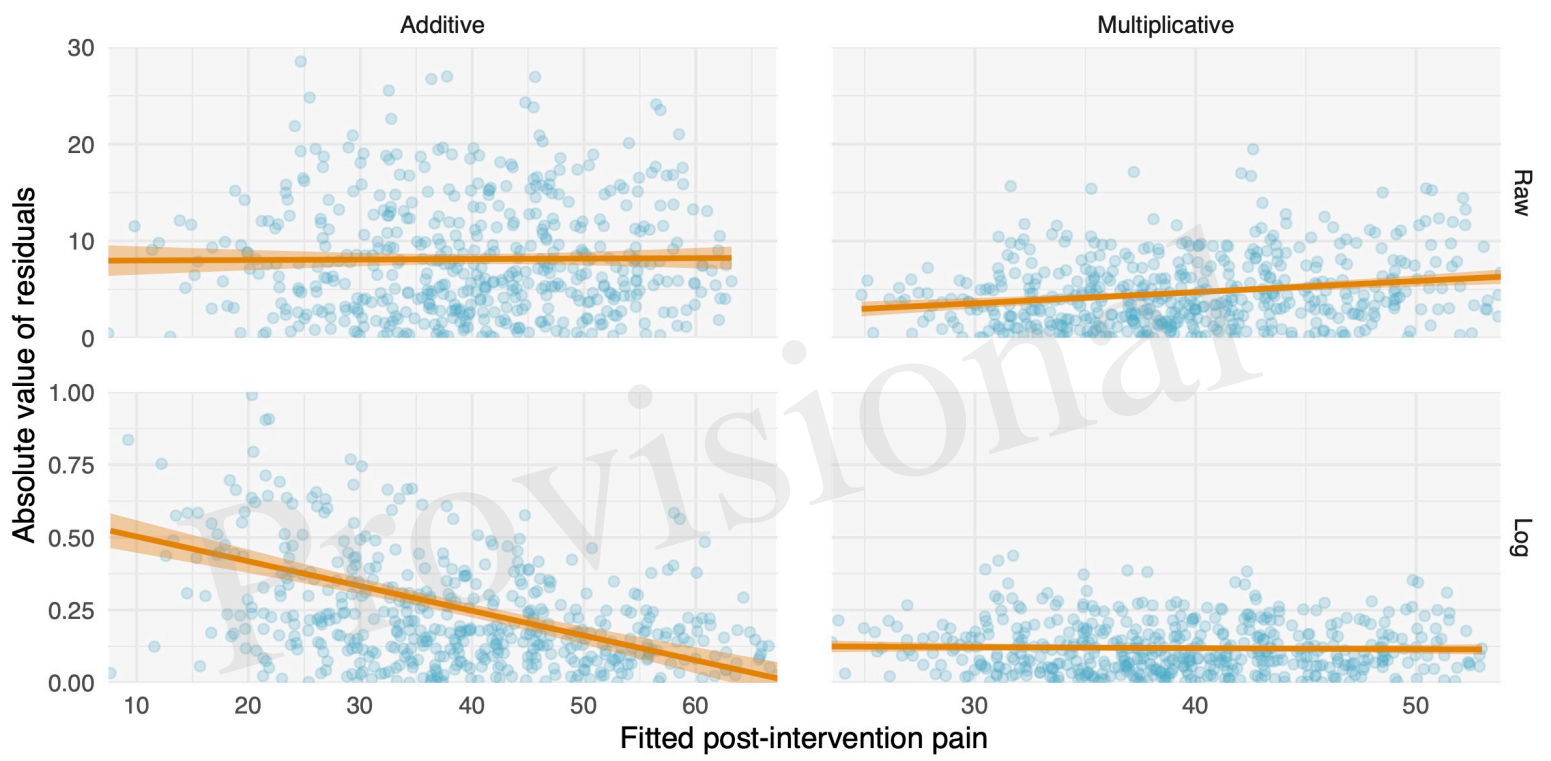


Figure 03.JPEG

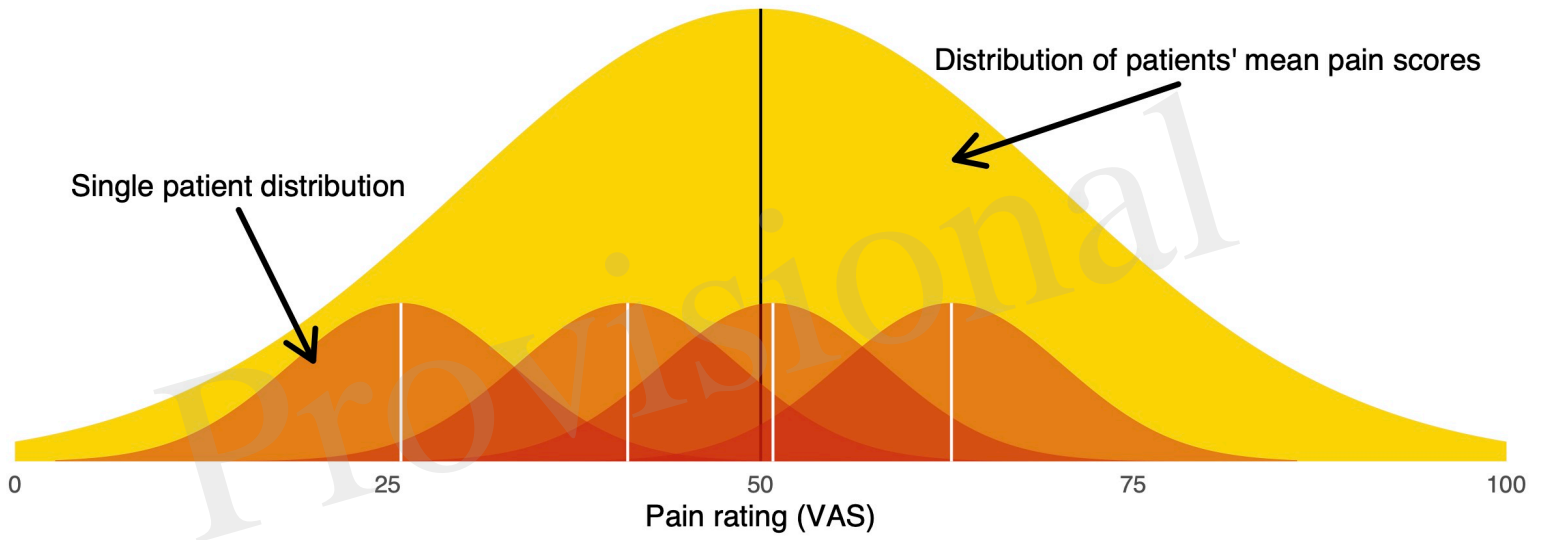


Figure 04.JPEG

Provisional

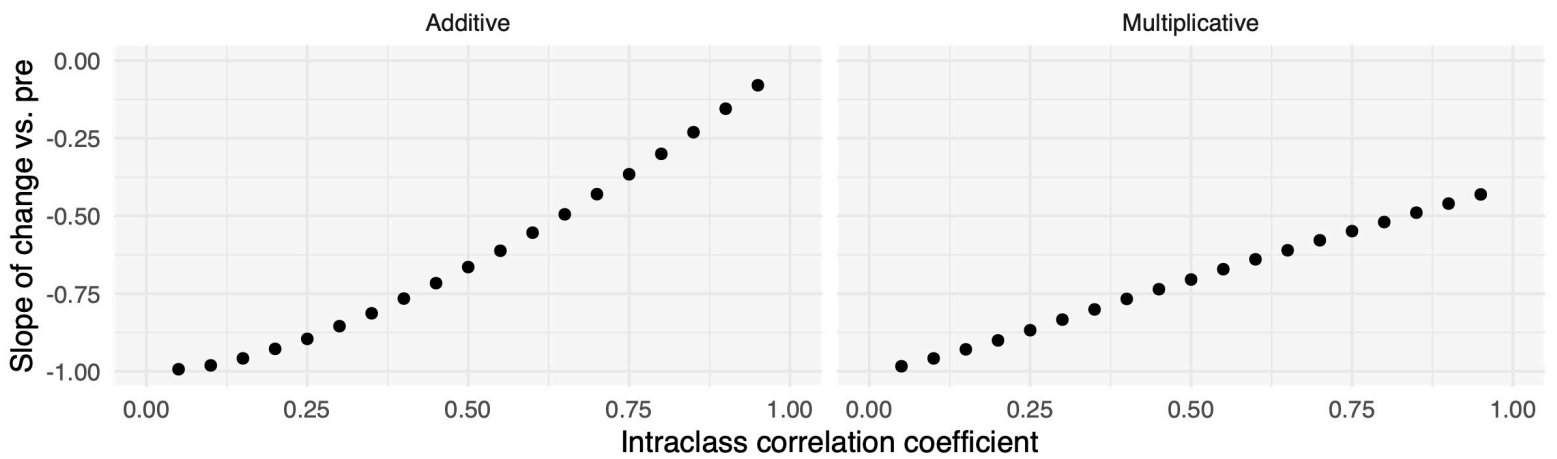


Figure 05.JPEG

Provisional

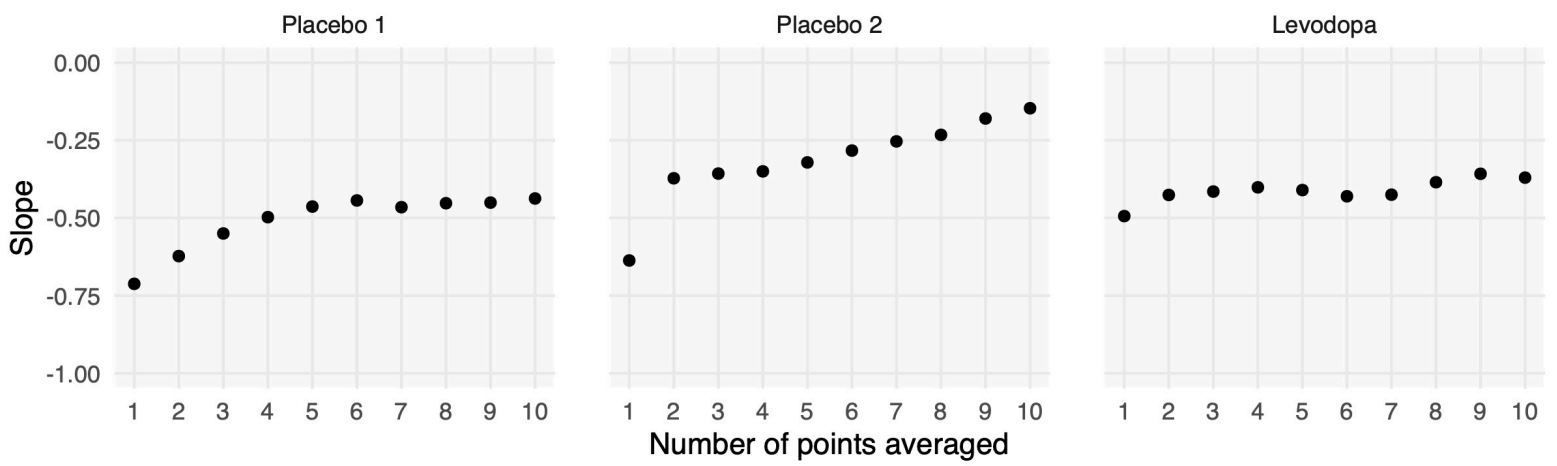


Figure 06.JPEG

