BMJ Open

Evaluation of person-level heterogeneity of treatment effects in published multi-person N-of-1 studies: systematic review and re-analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-017641
Article Type:	Research
Date Submitted by the Author:	08-May-2017
Complete List of Authors:	Raman, G; Tufts Medical Center Balk, EM; Brown University Lai, Lana; Tufts Medical Center Shi, Jennifer; Tufts Medical Center Chan, Jeffrey; VA Boston Healthcare System, Center for Healthcare Organization and Implementation Research (CHOIR) Lutz, Jennifer; Tufts Medical Center Dubois, Robert; National Pharmaceutical Council, Research Kravitz, Richard; University of California Davis Kent, David; Tufts Medical Center
Keywords:	perseonalized medicine, n-of-1 studies, systematic review, heterogeneity of treatment effect

Evaluation of person-level heterogeneity of treatment effects in published multi-person N-of-1 studies: systematic review and re-analysis

Gowri Raman, MD, MS[a]; Ethan M Balk, MD, MPH[b]; Lana Lai, MS[c]; Jennifer Shi, BA[d]; Jeff Chan, MD[a,e]; Jennifer Lutz, MA[c]; Robert Dubois, MD, PhD[f]; Richard L Kravitz, MD, MSPH[g]; David M Kent, MD, MS*[c]

[a]Center for Clinical Evidence Synthesis, ICRHPS, Medical Center/Tufts University School of Medicine, Boston, USA; [b]Center for Evidence Synthesis in Health, School of Public Health, Brown University, Providence, RI; [c]Predictive Analytics and Comparative Effectiveness Center, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center/Tufts University School of Medicine, Boston, USA; [d]Center for the Evaluation of Value and Risk in Health, ICRHPS, Medical Center/Tufts University School of Medicine, Boston, USA; [e] VA Boston Healthcare System, Center for Healthcare Organization and Implementation Research (CHOIR); [f]National Pharmaceutical Council; [g]Department of Internal Medicine, University of California, Davis

Corresponding Author:
David Kent, MD, MS
Director, Predictive Analytics and Comparative Effectiveness (PACE) Center
Institute for Clinical Research and Health Policy Studies
Tufts Medical Center
800 Washington St, Box 63
Boston, MA 02111
Email: dkent1@tuftsmedicalcenter.org
Phone: 617-636-3234

Running title: Variation in person-level treatment effects: systematic review
Word count
Abstract: 228
Main text: 3,970 (main text, references)
Table: 5
Figures: 3
Key words: n-of-1 studies, systematic review, heterogeneity of treatment effect, personalized medicine

Abstract

Objective: Individual patients with the same condition may respond differently to similar treatments. Our aim is to summarize the reporting of person-level heterogeneity of treatment effects (HTE) in multi-person N-of-1 studies and to examine the evidence for person-level HTE through re-analysis.

Study Design: Systematic review and re-analysis of multi-person N-of-1 studies.

Data sources: Medline, Cochrane Controlled Trials, EMBASE, Web of Science, and review of references through March 2014 for N-of-1 studies published in English.

Study Selection: N-of-1 studies of pharmacological interventions with at least two subjects.

Data Synthesis: Citation screening and data extractions were performed in duplicate. We performed statistical reanalysis testing for person-level HTE on all studies presenting person-level data.

Results: We identified 56 multi-person N-of-1 studies with at least two subjects. Statistical tests examining HTE were described in only nine (11%), of which only two (2%) tested person-level HTE. Only 23 studies (19%) in 24 data points provided person-level data sufficient to re-analyze person-level HTE. Reanalysis using a fixed effect linear model identified statistically significant person-level HTE in 8 of the 11 studies (73%) reporting person-level treatment effects and in 8 of the 13 studies (62%) reporting person-level outcomes.

Conclusions: Our analysis suggests person-level HTE is common and often substantial. Reviewed studies had incomplete information on person-level treatment effects and their variation. Improved assessment and reporting of person-level treatment effects in multi-person N-of-1 studies are needed.

Strengths and limitations of this study

- Multi-person N-of-1 studies are one of the best designs to estimate individual patient treatment effects and compare the variation in effects between individuals to variation within individuals across different periods
- This review highlights incomplete reporting of person-level treatment effects and their variation in multi-person N-of-1 studies.
- Re-analysis suggests person-level HTE is common and often substantial in multi-person N-of-1 studies.
- With improved assessment and reporting, multi-person N-of-1 studies have the potential to be important tools for personalized medicine.
- N-of-1 studies may be highly clinically informative for condition-treatments with a high degree of person-level HTE.

Introduction

Clinicians commonly observe that individual patients given the same treatment for the same condition frequently respond differently from one another. This observation, combined with our understanding of the complex mechanisms of diseases and therapies and the potential importance of myriad patient-specific factors (e.g., age, sex, illness severity, comorbidities, co-treatments, and molecular differences influencing pharmacokinetics and -dynamics), have led to a widely held assumption that the observed variation in treatment response seen between individuals is not merely random, but stable and potentially predictable. This assumption underpins the field of personalized medicine, which aims to determine the best treatment for an individual patient, as opposed to treating all patients with the same intervention found to be most effective for the "average" patient.

Nevertheless, statistical analyses aimed at discovering heterogeneity of treatment effects (HTE) among groups of individuals (for example subgroup analyses of parallel arm randomized trials) typically fail to find compelling and reliable evidence for the presence of such heterogeneity. Similarly, the field of pharmacogenetics, also built on the assumption of stable variation in treatment responses, has largely failed to live up to its promise to broadly improve the targeting of drugs—particularly outside the special case of oncology (where studies generally depend on the subclassification of tumor tissue not on variation in germline polymorphisms).[1,2] This failure to find reproducible HTE has supported the contrarian notion that true individual effects may be a "myth," an over-interpretation of random noise.[3]

To distinguish between these two possibilities, Kalow et al.[4] have suggested that carefully designed series of N-of-1 studies could be performed for those chronic conditions amenable to this design (i.e., where the disease process is relatively stable over time, treatment effects are

transient, and outcomes vary and are observable over time). By estimating individual patient treatment effects and comparing the variation in effects between individuals to variation within individuals across different periods, it is possible to determine heterogeneity in individual treatment effects--even if one is unable to identify the variables that predict this variation (i.e., even in the absence of group-level HTE, such as men versus women, or old versus young).

A recent review[5] summarized N-of-1 studies reported in the literature—including multi-person N-of-1 studies—but did not examine whether and how these studies provide information on person-level HTE. Therefore our objectives are: 1) to summarize the conduct and reporting of assessments of variation in person-level treatment effects from N-of-1 studies; and 2) to extract, reanalyze and report the results from the subset of studies that provided adequate data in their published reports to examine the extent of the evidence for person-level HTE (i.e., participant-level outcomes or effects).

Methods

This review was conducted in accordance with the highest standards for conducing systematic reviews.[6,7] We defined N-of-1 studies as crossover trials in which each patient receives two or more treatments in a pre-defined, often randomized, sequence.

Data Sources and Searches

We used two separate searches because N-of-1 studies can be indexed differently: (1) a search in Medline, Cochrane Central and EMBASE using terms related to repeated crossover studies (for publications indexed from inception to March 21, 2014); and (2) a Medline, Cochrane Central, EMBASE, and Web of Science search using terms that are related to N-of-1

(for publications indexed from 2011 to March 21, 2014). For N-of-1 studies indexed before 2011, we used studies included in a prior published systematic review by Gabler et al.[5] Our searches combined terms and Medical Subject Headings for N-of-1, single-subject, single-patient, randomized trials, crossover, multi-period crossover, and rotated or repeated period crossover (see Appendix Tables 1-2 for detailed search terms). The searches were not restricted by disease, condition, organ system, or treatment.

Study Selection

We selected eligible multi-person N-of-1 studies to describe the frequency of reporting of individual outcomes and effects and of documented HTE in these studies. We required that a minimum of two individual subjects per study for evaluation of HTE. We excluded studies that included non-pharmacological interventions, reviews, abstracts and protocols. We include studies with placebo or "no treatment" interventions. Citations were double-screened by reviewers using an open-source, online software Abstrackr (http://abstrackr.cebm.brown.edu/). Full-text articles of potentially relevant studies were again double screened for eligibility.

Person-level outcomes were defined as outcomes for each person at each point in time when they were measured, reported in tables, text, or graphs. Person-level treatment effect was defined as contrasts of outcomes in individuals on one treatment versus the comparator. Person-level HTE was defined as quantified variation in the person-level treatment effects, whereas HTE more broadly includes any type of subgroup analysis (e.g., males versus females; older versus younger) as outlined in **Figure 1**.

Data Extraction and Quality Assessment

One of four reviewers extracted data from each publication; a second reviewer verified

all numerical information and basic descriptors of the study design and analysis. Operational definitions for extraction items were discussed in weekly project meetings and discrepancies between extractors were resolved by consensus with senior authors (DK, GR, EB). From each study, we extracted bibliographic information, details related to study design (number of patients enrolled, selection criteria, interventions evaluated, randomization methods, outcomes assessed, follow-up duration), information on patient characteristics, and person-level measurements of outcomes or estimates of person-level treatment effects (with corresponding measures of their uncertainty). When necessary, we extracted data by digitizing the graphs and the values were estimated using Engauge Digitizer version 2.14 (http://digitizer.sourceforge.net/).

We generated graphs showing the trajectory of response for each patient in each study and compared them against the published information. We also generated scatterplots of measurements over time for studies that did not present their data in graphical format to help us identify aberrant data points (e.g., errors in data extraction). We verified potentially aberrant data points by re-examining the published data and made corrections, when needed.

Data Synthesis and Analyses

We examined the degree to which studies reported person-level data. This was described using the following items for each reported outcome: 1) qualitative descriptions of HTE (e.g., "there were 8 responders and 4 non-responders"); 2) details of person-level outcomes (i.e., outcomes with each treatment within each period); 3) details of person-level treatment effect (i.e., a point estimate of contrasts of outcomes in individuals on one treatment versus the comparator); 4) reporting of person-level statistical effect estimate, (e.g., standard deviation, exact P values, or confidence intervals for treatment effects within individuals); 5) description of statistical tests examining HTE (i.e., tests evaluating the contrast of treatment effects between

individuals or groups in the study); and 6) claims of HTE. Note that qualitative descriptions of HTE for item 1 would include any description that implied that treatment effects varied, whereas item 6 required a more definite study conclusion (e.g., "our results demonstrate significant variation across individuals in response to treatment X"), whether or not these conclusions were based on robust statistical tests.

Statistical HTE analysis of extracted study results

We performed statistical analysis testing for person-level HTE on all studies presenting person-level data. We used a consistent analytic strategy across studies, to the extent permitted by the reporting in published papers. Our strategy was different for studies that reported person-level outcome measurements and those that reported estimates of person-level treatment effects with their sampling variances (or adequate information to approximately calculate these statistics).

For studies that only reported (or allowed the calculation of) *estimates of person-level treatment effects*, we obtained an average effect using a fixed effect inverse variance model and estimated the variance of the person-level treatment effects using a method of moments estimator. In addition to a fixed effect model, we also obtained an average effect using a random effects model. Finally, we tested the hypothesis that all person-level treatment effects were equal using Cochran's chi-square test and quantified the proportion of observed variation due to "true" person-level effect heterogeneity with the I^2 statistic.[8]

For studies that reported *person-level outcomes*, we developed a linear model (for continuous outcomes) or generalized linear model (for binary or count outcomes) using the outcome of interest as the response, the intervention(s) as a covariate; indicator variables for different study participants were derived. This model estimates a common treatment effect across

participants. We also derived a similar model with treatment-by-participant interactions. This model allows each patient to have a different effect. The statistical significance of person-level HTE was assessed by a likelihood ratio test comparing the two models. In addition to a fixed effect model, we also fit a hierarchical linear or generalized linear mixed model with a random intercept and a random slope (for the treatment effect) to estimate the average treatment effect across all patients (assuming person-level HTE). We tested the hypothesis that all person-level treatment effects were equal and quantified the proportion of observed variation due to 'true' person-level effect heterogeneity with the I^2 statistic.[8]

Results

The searches for repeated crossover studies identified 10,596 citations and those for N-of-1 studies identified 2676 citations (indexed from 2011 onwards). Of these, we retrieved 373 full-text articles for review plus 100 N-of-1 trial articles (indexed before 2011) from an existing systematic review.[5] Upon full-text screening, 56 studies (52 multi-person N-of-1 studies and four repeated period crossover studies) met eligibility criteria (Appendix Table 3) and are reported multi-person N-of-1 studies throughout the article. An outline of the search and study selection flow is provided in **Figure 2**.

Description of studies

Table 1 summarizes the 56 multi-person N-of-1 studies that were published between 1986 and 2014 reporting a total of 1974 patients. The most common clinical domains in the multi-person N-of-1 studies were neurology (23%), arthritis/rheumatology (18%) and psychiatry (16%). Most studies were described as "double-blind" but details about the methods for blinding

were often unclear; similarly studies often provided unclear information about the generation of the randomization sequence and allocation concealment. Among the studies, 93% compared a pair of treatment strategies, 5% compared three strategies, and 2% compared four strategies. Studies had between 3 and 16 treatment periods and obtained an average of 1 to 42 outcome measurements per period. Across reported outcomes, 89% of the assessed outcomes were patient-reported and 11% were investigator-assessed.

Reporting Person-level outcomes, effects and HTE

While most studies (95%) had some qualitative acknowledgement that the treatment effects appeared to vary across individuals, formal reporting at the participant level was variable **(Table 2)**. Person-level outcomes under each treatment were reported in 48% of multi-person N-of-1 studies. Person-level treatment effects with quantitative data (comparing outcomes on each treatment) for each individual who completed the trial was available in 29%; and details on the statistical evaluation of these effects (as standard deviations or exact P values or confidence intervals) were available in only two multi-person N-of-1 study. Only four (7%) studies described statistical tests examining any HTE. However, only two studies reported person-level HTE, whereas the other two examined group-level HTE using conventional subgroups.

Reanalysis of person-level data:

Of the 56 studies, there were 31 studies that provided person-level data, either as outcomes in each treatment period or as person-level treatment effects **(Table 3)**. Of these, only 23 studies provided person-level data sufficient to support re-analysis: 13 studies provided person-level outcomes; 11 studies provided person-level treatment effects (one study provided

both). The remaining eight studies reported either medians or means without data on variance, so they could not be re-analyzed for treatment effect or HTE.

Of 11 studies (with 19 unique comparisons) that reported analyzable person-level treatment effect data **(Table 3)**, seven studies had a placebo comparator and three studies compared had an active comparator. The sample size ranged from 7 to 68; average crossover periods ranged from 6 to 16 days; and average outcome measures per period ranged from 1 to 21. The average treatment duration ranged from 14 to 336 days.

There were 13 studies (with 26 unique comparisons) that reported analyzable person-level outcome data **(Table 3)**, including one study also reporting person-level outcomes. Of these, 10 compared the intervention with placebo and three studies compared two active interventions. The sample size ranged from 2 to 22; the average number of crossover periods ranged from 3 to 10; and the average number of outcome measures per period ranged from 1 to 42. The average treatment duration ranged from 9 to 210 days.

Re-analysis of studies reporting estimates of person-level treatment effects

Eleven studies (including 19 comparisons, due to multiple outcomes in some studies) reported estimates of person-level treatment effects sufficient to analyze (Appendix Figures 1-14 displays graphs of the person level treatment effect data). Average fixed effect estimates for each analysis are shown in **Table 4**; random effects estimates were generally similar (Appendix 4). In 8 of the 11 studies (73%) and 14 of the 19 total unique comparisons (74%) we found evidence of statistically significant HTE for at least one outcome (Table 4). Generally, the magnitude in the variation of individual patient effects (as seen in the range) was very large compared to the average effects. Most studies (64%) showed person-level effects that differed qualitatively from

one another. Most of the variation in the observed individual effects was attributable to "true" heterogeneity of person-level effects; 11 of 19 analyses had I^2 >80%.

Re-analysis of studies reporting person-level outcome measurements

Because some of the 13 studies providing analyzable outcome data had multiple outcomes (or multiple outcomes scales) there were a total of 26 comparisons with analyzable data. (Appendix Figures 15-40 displays graphs of the person level outcome results.) Average fixed effect estimates for each analysis are shown in Table 5; random effects estimates were generally similar (Appendix 5). In eight of the 13 studies (62%) (17 of the 26 unique comparisons [65%]), there was statistically significant person-level HTE for at least one outcome. Among the 26 unique outcome comparisons in eight studies, 17 outcomes (65%) demonstrated statistically significant person-level HTE. Again, the variation in individual effects was often large compared to the average effect. However, given the lower number of participants per study and periods per participant and also different analytic approach, estimates of I^2 were much less precise in these studies.

Discussion

This review documents that multi-person N-of-1 studies rarely examine HTE. Only 8% of 56 multi-person N-of-1 studies described statistical tests examining HTE, but these generally involved comparisons of treatment effects among groups of patients (e.g., based on age or sex) rather than across individuals. Only two studies in the whole of the literature tested for person-level THE.[9,10] Nevertheless, analyzable person-level results are sometimes reported in multi-person N-of-1 studies, as outcomes or as treatment effects. Our re-analyses of the totality of available data from these studies (n=31) suggested the presence of substantial variation in treatment effects across individuals in most studies. This was evident when considering

statistical tests for the variation of treatment effects among patients and also by qualitative assessment of the magnitude of effect variation. This represents the first broad empirical examination with re-analysis of person-level HTE across multi-person N-of-1 studies, and it provides some general support for the *a priori* assumption of individual patient variation in treatment response that broadly motivates personalized medicine.

In contrast to parallel-group studies that establish efficacy in a group of patients with a common condition, N-of-1 studies establish the effects of an intervention in an individual.[11] In this respect, N-of-1 studies can be thought of as adjuncts to clinical care, where the goal is to select the right treatment for a particular patient, rather than as a research tool, where the goal is to create new generalizable knowledge.[12,13] Indeed, the results of traditional N-of-1 studies may be generalizable only to the future treatment response of the patient in the trial; it may not apply to other patients. Nevertheless, using Bayesian meta-analytic techniques, Zucker et al. showed how the average treatment effect at the population-level can also be estimated from combining multi-person N-of-1 studies testing similar interventions in similar patients with the same outcome measures.[14]

Herein, we demonstrate yet a new application of N-of-1 studies, to explore person-level HTE to describe the variation in individual treatment effects. This application has important research and clinical implications, even when the determinants of HTE remain unidentified. It is particularly of interest that there was apparent variation in the *degree* of person-level HTE found across conditions and treatments. Since the degree of variation across individuals sets the upper bound for the amount of HTE that might be explainable by observable characteristics, such as clinical or genomic variables, searching for subgroup effects in the absence of person-level HTE is a futile exercise.[3,15]

An interesting example of how person-level HTE can vary across different conditions comes from the study of Johannessen et al **(Figure 3)**.[9] These investigators conducted N-of-1 patient studies comparing cimetidine to placebo for patients presenting with dyspeptic symptoms and reported person-level effects by subgroups of disease categories. Among 46 trial completers, cimetidine had a significant effect for most patients (57%) and at the aggregate level. However, not only was there substantial person-level HTE, but person-level HTE varied across conditions, being much more pronounced in non-ulcer dyspepsia (I^2 = 75%) compared to peptic ulcer disease (I^2 = 35%) (Figure 3)— despite the very similar overall effects seen in these two conditions.

Finding variation in person-level response in multi-person N-of-1 studies identifies those conditions for which N-of-1 studies are likely to be clinically relevant. For condition-treatment combinations shown to have low person-level HTE, single subject studies are highly unlikely to be clinically informative, and the average results from trials (i.e., "one-size-fits-all" effects) are more apt to be applicable to individuals.[16,17] On the other hand, N-of-1 studies may be highly clinically informative for condition-treatments with a high degree of person-level HTE. These conditions would also be potentially higher yield for examining predictors of HTE (genomic or otherwise).

Our findings also have implications for clinical practice and formulary design. For conditions marked by high person-level HTE, even when trials show that one treatment is better on average than others, having a variety of medication options would be useful to optimize outcomes across all patients, particularly for chronic conditions such as those studied here where empiric trials of alternative medications to find the best treatment for an individual might be feasible. For example, the study by March et al.[18] shows that while patients with osteoarthritis on

average had less pain and less stiffness with diclofenac, some patients had improved symptoms on paracetemol. This person-level heterogeneity of treatment effect may not be detectable in conventional parallel arm trials employing conventional subgroup analysis.[15]

While more studies combining N-of-1 studies are needed to understand the extent of person-level HTE, future studies need to apply greater methodological rigor to improve the state-of-the-science on evaluation of individual treatment effects.[19] While the recently published CONSORT Extension for N-of-1 Trials[20] may help improve reporting, given the relatively small number of individuals enrolled in each study and the relatively few treatment periods that are typical in medical studies, a tabulation of all information (possibly electronically available) appears the most straightforward way to facilitate the clinical interpretation of these studies. Such reporting allows the inspection of trajectories over time and may reveal patterns that are not captured by regression models. Complete reporting would also facilitate the development and evaluation of methods for the analysis of single subject experiments.

The limitations of this review reflect, to a large extent, the limitations of the data in primary studies. Many important disease categories lacked published N-of-1 studies, even though potentially amenable to this design. We relied on published studies only and our analytic cohort may be an underestimation of the true prevalence of these studies—particularly for N-of-1studies, which may frequently be conducted without the intention of future publication.

In addition, our conclusions regarding the ubiquity of HTE in the data we reanalyzed should be interpreted in the context of several important limitations. First, there were only a limited number of available studies that reported data sufficient to analyze, and therefore we present only a very partial picture of the full scope of inter-individual variation in effects across clinical conditions. Furthermore, among the studies that did have data, only fairly small numbers

of patients were observed over a small number of treatment periods and we frequently had to rely on data summaries provided by the authors (e.g., person-level treatment effects and their sampling variance); these data limitations precluded the use of more complex models, for example models that account for period effects or other effects of time on the outcome.[3]

Our review has demonstrated that HTE remains almost totally unexplored in multi-person N-of-1 studies, which are uniquely capable of exploring variations in individual (person-level) treatment effects. Our re-analysis of the data from these studies represents the first systematic attempt to obtain empirical support for the *a priori* argument that treatment effects vary across individual patients, an assumption which underpins all efforts to personalize treatment selection. In this sample, person-level HTE appears to be fairly common and large enough to be clinically meaningful; the degree of person-level HTE appears to vary across conditions and outcomes. Thus, multi-person N-of-1 studies are an under-utilized tool to identify where person-level HTE may be substantial, and where efforts to find molecular or clinical predictors of response heterogeneity should be focused. In such conditions, parallel arm studies might yield results that are over-generalized for patient level decision making.

Funding Source

This work was supported by the National Pharmaceutical Council.

Acknowledgement

We would like to acknowledge Issa Dahabreh, MD, MS, Assistant Professor of Health Services, Policy and Practice, Assistant Professor of Epidemiology, Brown University, for statistical advice.

Contributorship statement

DK, GR made substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work. DK, GR are responsible for drafting the work or revising it critically for important intellectual content. DK, GR, EB, LL, JS, JC, JL, RD, RK have given final approval of the version to be published. DK, GR have made an agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Competing interests

None declared.

Data sharing statement

No additional data are available.

.

References

1. Hertz DL, McLeod HL. Use of pharmacogenetics for predicting cancer prognosis and treatment exposure, response and toxicity. *J Hum Genet.* 2013;58(6):346-352.
2. Kitsios GD, Kent DM. Personalised medicine: not just in our genes. *BMJ.* 2012;344:e2161.
3. Senn S. Individual response to treatment: is it a valid assumption? *BMJ.* 2004;329(7472):966-968.
4. Kalow W, Tang BK, Endrenyi L. Hypothesis: comparisons of inter- and intra-individual variations can substitute for twin studies in drug research. *Pharmacogenetics.* 1998;8(4):283-289.
5. Gabler NB, Duan N, Vohra S, Kravitz RL. N-of-1 trials in the medical literature: a systematic review. *Med Care.* 2011;49(8):761-768.
6. Research IoMUCoSfSRoCE. Finding What Works in Health Care: Standards for Systematic Reviews. 2011.
7. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ.* 2009;339:b2535.
8. WG C. The combination of estimates from different experiments. *Biometrics.* 1954;10:101-129.
9. Johannessen T, Petersen H, Kristensen P, et al. Cimetidine on-demand in dyspepsia. Experience with randomized controlled single-subject trials. *Scand J Gastroenterol.* 1992;27(3):189-195.
10. Pereira JA, Holbrook AM, Dolovich L, et al. Are brand-name and generic warfarin interchangeable? Multiple n-of-1 randomized, crossover trials. *Ann Pharmacother.* 2005;39(7-8):1188-1193.
11. Guyatt GH, Heyting A, Jaeschke R, Keller J, Adachi JD, Roberts RS. N of 1 randomized trials for investigating new drugs. *Control Clin Trials.* 1990;11(2):88-100.
12. Guyatt G, Sackett D, Taylor DW, Chong J, Roberts R, Pugsley S. Determining optimal therapy--randomized trials in individual patients. *N Engl J Med.* 1986;314(14):889-892.
13. Guyatt G, Sackett D, Adachi J, et al. A clinician's guide for conducting randomized trials in individual patients. *CMAJ.* 1988;139(6):497-503.
14. Zucker DR, Ruthazer R, Schmid CH. Individual (N-of-1) trials can be combined to give population comparative treatment effect estimates: methodologic considerations. *J Clin Epidemiol.* 2010;63(12):1312-1323.
15. Dahabreh IJ, Hayward R, Kent DM. Using group data to treat individuals: understanding heterogeneous treatment effects in the age of precision medicine and patient-centred evidence. *Int J Epidemiol.* 2016. [Epub ahead of print]
16. Simon G. Choosing a first-line antidepressant: equal on average does not mean equal for everyone. *JAMA.* 2001;286(23):3003-3004.
17. Simon GE, Psaty BM, Hrachovec JB, Mora M. Principles for evidence-based drug formulary policy. *J Gen Intern Med.* 2005;20(10):964-968.
18. March L, Irwig L, Schwarz J, Simpson J, Chock C, Brooks P. n of 1 trials comparing a non-steroidal anti-inflammatory drug with paracetamol in osteoarthritis. *BMJ.* 1994;309(6961):1041-1045.

19. RL K, N D, Panel DMCN-o-G. Design and Implementation of N-of-1 Trials: A User's Guide. AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2014.
20. Vohra S, Shamseer L, Sampson M, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. *J Clin Epidemiol.* 2016;76:9-17.

Table 1. Evidence Map of Multi-person N-of-1and Repeated Period Crossover Studies

Description	Multi-person N-of-1 Studies (n=56)
Publication Years	1979-2014
Subjects	**Total N (median, IQR)**
Enrolled	1974 (21, 10-43)
Completed	1573 (14, 14-34)
Intervention & Comparisons	
Head-to-head active drugs	10
Placebo	41
Active drug and placebo	1
Population	
Pediatric	11
Adults	41
Major Systems Studied	
Arthritis/Rheumatology	10
Cardiovascular	3
Gastrointestinal	7
Hypertension	0
Psychiatry	9
Neurology	13
Respiratory	7
Miscellaneous*	7
Top 5 Disease Conditions	
ADHD	6
Angina	3
Chronic Pain	5
GERD	5
Obstructive Airway	6
Osteoarthritis	6

**Sleep disorders, Allergy, Cancer, Muscular, Vascular (for multi-person N-of-1); Pain, Urology, GYN, Rheumatology, Heme/Onc, Allergy, Dermatology, Drug abuse, Endocrine, Lipids, Nephrology, Ophthalmology, Respiratory (for Repeated Cross-over Studies).*
ADHD, Attention-deficit hyperactivity disorder; GERD, Gastro-esophageal regurgitation disorder; IQR, Interquartile range; n, number of participants

Table 2. Survey of HTE Assessment in Multi-person N-of-1 Studies

HTE Reporting	Multi-person N-of-1 Studies (n=56)
Qualitative description	95%
Person-level outcomes	48%
Person-level treatment effects	29%
Statistical analysis of person-level effects (e.g. p-values)	4%
Any statistical test for HTE	7%*
Claims of heterogeneity	10%

* Only 2 studies reported person-level HTE, the remaining 2 studies reported group level effect.

Table 3. Characteristics of studies reporting person-level data

Author, Year	Disease	Number enrolled (analyzed)	Intervention	Comparator	Cross-over periods	Total intervention duration	Outcome measures per period
Studies with re-analyzable person-level outcomes							
Camfield, 1996	Mental retardation with fragmented sleep	6 (6)	Melatonin	Placebo	7	10 wk	14
Hinderer, 1990	Traumatic spinal cord injury	5 (5)	Baclofen	Placebo	3	9 wk	2
Langer, 1993	Gastroesophageal reflux	2 (2)	Cisapride	Placebo	3	6 wk	5
Lashner, 1990	Ulcerative colitis	7 (6)	Nicotine	Placebo	4	8 wk	1
Maier, 1994	Chronic depression	10 (9)	Sulpiride	Placebo	4	28 wk	42
Mandelcorn, 2004	Brain injury	4 (4)	Ondansetron	Placebo	4	5 wk	1
McQuay, 1994	Neuropathic pain	19 (19)	Dextromethorphan	Placebo	5	20 d	1
Miyazaki, 1995	Unstable angina	22 (22)	Isosorbide dinitrate	Isosorbide dinitrate: intermittent injection	3	9 d	6
Nathan, 2006	Pediatric brain tumor	12 (7)	Ondansetron & metopimazine	Ondansetron & placebo	Unclear	189 d	unclear
Parodi, 1979	Unstable angina	12 (12)	Verapamil	Placebo	4	10 d	unclear
Parodi, 1986	Unstable angina	10 (10)	Verapamil	Propranolol, placebo	8	18 d	unclear
Tison, 2012	Levodopa-induced dyskinesia in Parkinson's disease patients	10 (10)	Simvastatin	Placebo	6	96 d	1
Studies with re-analyzable person-level treatment effects							
Emmanuel, 2012	Chronic intestinal pseudo-obstruction	7 (4)	Prucalopride	Placebo	16	48 wk	21
Haas, 2004	Chronic tension-type and migraine headache	39 (16)	Dextroamphetamine	Equi-stimulatory caffeine	8	20 d	20
Jaeschke, 1991	Fibromyalgia	22 (23)	Amitriptyline	Placebo	6	12 wk	2
Johannessen, 1992	Dyspepsia	68 (46)	Cimetidine	Placebo	12	184 d	15
Mahon, 1996	Irreversible chronic airflow	16 (14)	Theophylline	Placebo	8	73 d	1

	limitation						
March, 1999	Osteoarthritis	25 (15)	Diclofenac	Paracetamol	6	12 wk	14
Patel, 1991	Nonreversible chronic airflow limitation	26 (18)	Ipratropium bromide / theophylline / salbutamol/ beclomethasone	Placebo	6	6 wk	Unclear
Wallace, 1994	Attention deficit hyperactivity disorder	11 (7)	Methylphenidate	Placebo	14	14 d	1
Woodfield, 2005	Skeletal muscle cramps	13	Quinine	Placebo	6	14 wk	2
Zucker, 2006	Fibromyalgia	58	Amitriptyline and Placebo	Amitriptyline and fluoxetine combination	6	36 wk	1
Study with both person-level data							
Pereira, 1995	Atrial fibrillation / deep venous thrombosis	7	Generic warfarin	Coumadin	10	30 wk	2
Study with insufficiently reported person-level data							
Person-level outcome data							
Denburg, 1994	Systemic lupus erythematosus	10	Prednisone	Placebo	6	30 wk	1
Nikles, 2000	Osteoarthritis	14	Ibuprofen	Paracetamol; Placebo	6	12 wk	14
Reitberg, 2002	Allergic rhinitis	36	Loratadine and chlorpheniramine maleate	loratadine with placebo	8	32 d	4
Sheather-Reid, 1998	Chronic pain	8	Ibuprofen / Codeine	Placebo	6	12 wk	14
Person-level treatment effects							
Huber, 2007	Juvenile idiopathic arthritis	6	Amitriptyline	Placebo	6	17 wk	12
Privitera, 1994	Partial seizure	16	Dezinamide	Placebo	6	35 wk	6
Wegman, 2003	Osteoarthritis	13	Paracetamol	NSAIDs	10	20 wk	14
Wegman, 2005	Regular Temazepam users	15	Temazepam	Placebo	10	10 wk	7

Table 4. Analysis results of studies reporting person-level treatment effects

			Main Effect	*Person-Level Heterogeneity of Treatment Effect*		
Author Year	**Outcome**	**Range of the scales (severity)**	**Treatment effect (CI)**	**P for HTE***	**Treatment Effect Range Lower Range (CI) Upper Range (CI)**	**I-square % (CI)**
Emmanuel 2012	Bloating	0-4 (0=absent to 4=worst)	-0.344 (-0.619 to -0.069)	<0.001	-1.1 (-1.37 to -0.83) -0.1 (-0.27 to 0.07)	94 (88 to 97)
	Pain	0-4 (0=absent to 4=worst)	-0.440 (-0.771 to -0.110)	<0.001	-0.2 (-0.33 to -0.77) -1.4 (-1.69 to -1.11)	96 (92 to 98)
Haas 2004	Chronic tension-type headache grade	0-3 (0=none to 3=severe)	0.772 (0.454 to 1.090)	<0.001	0.04 (-0.39 to 0.47) 1.9 (1.29 to 2.50)	84 (76 to 90)
	Chronic migraine headache grade	0-3 (0=none to 3=severe)	0.542 (0.354 to 0.731)	0.067	0.2 (-0.41 to 0.81) 0.83 (0.24 to 1.42)	37 (0 to 65)
Jaeschke 1991	7-point symptom scale	1-7 (higher scores represent better function)	0.427 (0.210 to 0.645)	<0.001	-1.02 (-2.82 to 0.77) 3.18 (1.89 to 4.46)	85 (79 to 89)
	Tender point changes count	Number of tender points	1.320 (0.404 to 2.236)	<0.001	-4.33 (-10.8 to 2.14) 9.0 (5.36 to 12.63)	72 (57 to 82)
Johannessen 1992	6-point symptom scale	0-6 (0=NR to 6=NR)	0.698 (0.466 to 0.931)	<0.001	-1.67 (-2.78 to -0.55) 3.17 (0.60 to 5.74)	66 (53 to 75)
Mahon 1996	Dyspnea in likert Scale	1-7 (1=extremely short of breath to 7=no shortness)	0.125 (-0.181 to 0.430)	<0.001	-0.57 (-1.55 to 0.42) 0.89 (0.62 to 1.16)	78 (58 to 88)
March 1999	Mean pain score on VAS	5 point Likert scale (0-100mm)	-7.093 (-11.939 to -2.248)	<0.001	-33.8 (-38.9 to -28.6) 4.1 (-17.0 to 25.0)	98 (97 to 98)
	Mean stiffness score on VAS	5 point Likert scale (0-100mm)	-5.992 (-11.280 to -0.704)	<0.001	-36 (-50.6 to -21.4) 10.7 (1.12 to 20.2)	97 (96 to 98)
Patel 1991**	4-item symptom questionnaire (All compared to placebo)	1-7 (1=extremely short of breath to 7=no shortness of breath)	0.240 (0.131 to 0.350)	<0.001	-0.34 (-1.04 to 0.36) 3.1 (1.54 to 4.66)	91 (87 to 94)
	4-item symptom questionnaire (use of ipratropium bromide)		0.675 (0.264 to 1.085)	<0.001	-0.22 (-0.71 to 0.26) 3.1 (1.54 to 4.66)	87 (78 to 92)
	4-item symptom questionnaire (use of salbutamol)		0.865 (0.042 to 1.687)	<0.001	0.46 (0.27 to 0.65) 1.3 (0.93 to 1.67)	94 (NA)
	4-item symptom questionnaire (use of theophylline)		0.025 (-0.434 to 0.484)	0.172	-0.34 (-1.04 to 0.36)	30 (0 to 93)
Pereira 1995	INR (diff)	Target INR range of 2.0–3.0	0.027 (-0.155 to 0.209)	0.477	-0.28 (-0.97 to 0.41) 0.37 (-0.07 to 0.81)	0 (0 to 75)
Wallace 1994	Conners 15-item rating scale scores	0-3 (NR)	0.759 (0.341 to 1.178)	0.747	0.42 (-0.51 to 1.35) 1.22 (0.23 to 2.21)	0 (0 to 79)
Woodfield 2005	Changes in number of cramps	Number – mean difference	-18.823 (-28.527 to -9.120)	<0.001	-77 (-106.0 to -47.9) -2 (-3.96 to -0.04)	92 (87 to 95)
	Total days with cramps	days	-6.181 (-9.798 to -2.563)	<0.001	-13 (-14.98 to -11.02) -1 (-6.15 to 4.15)	94 (90 to 96)
Zucker 2006	FIQ	0-100 (0=best to 100=worst)	-5.019 (-8.784 to -1.254)	0.999	-32.0 (-79.2 to 15.18) 0.98 (-16.9 to 18.9)	0 (0 to 37)

* The significance of person-level HTE was assessed by Cochran's chi-square-based test
** One subject had beclomethasone

Table 5. Studies reporting person-level outcomes

Author Year	Outcome	Definition / Range of the Scales (severity)	*Main Effect*	*Person-level Heterogeneity of Treatment Effect*		
			Fixed Treatment Effect	P for Person Treatment Interaction*	Treatment Effect Range Lower Range (CI) Upper Range (CI)	I-square % (CI)
Camfield 1996	Nights without awakening	Between 10:00 PM and 7:00 AM per day	0.865 (0.215 to 1.516)	0.456	0.12 (-1.9 to 2.2) 2.0 (-0.1 to 4.2)	0 (0 to 79)
Hinderer 1990	Anxiety	Beck Inventory-A anxiety scale 0-3 (0 = never, 3 = almost all the time)	0.000 (0.000 to 0.000)	<0.001	-6.38 (-7.96 to -4.80) 0.000 (-1.64 to 1.64)	91 (81 to 95)
Langer 1993	Vomiting	Number of episodes	-1.204 (-2.494 to 0.086)	0.136	-1.34 (-0.76 to 0.22) 0.17 (-0.41 to 0.76)	87 (NA)*
Lashner 1990	Symptom score: abdominal pain	Symptom scores 0-100 (0 = best, 100 = worst)	-3.615 (-16.982 to 9.751)	0.007	-35.0 (-65.8 to -4.2) 15.0 (-28.6 to 58.6)	37 (0 to 73)
	Symptom score: bowel movements/day		-0.538 (-1.215 to 0.138)	0.001	-3.0 (-4.9 to -1.0) 1.0 (-0.4 to 2.4)	56.6 (0 to 81)
	Symptom score: consistency of bowel movements		7.000 (-7.551 to 21.551)	0.013	-25.5 (-60.1 to 9.1) 33.0 (-1.6 to 67.6)	28 (0 to 69)
	Symptom score: hematochezia		2.308 (-17.210 to 21.826)	0.003	-38.0 (-80.8 to 4.8) 47.5 (4.7 to 90.3)	47 (0 to 78)
	Symptom score: general sense of well-being		-6.538 (-25.352 to 12.275)	0.008	-43.0 (-104.6 to 18.6) 35.0 (-8.6 to 78.6)	35 (0 to 73)
Maier 1994	SCL-90 subscales: Depressed mood	Self-rating inventory to measure the effects of drug	-3.536 (-6.718 to -0.354)	<0.001	-17.8 (-25.5 to -10.1) 2.74 (-4.9 to 10.4)	58 (12 to 80)
	SCL-90 subscales: Anxiety		-3.753 (-6.582 to -0.924)	<0.001	-17.4 (-23.8 to -10.9) 2.5 (-3.9 to 8.9)	66 (30 to 83)
	SCL-90 subscales: Somatization		-1.419 (-4.316 to 1.478)	0.869	-6.0 (-16.0 to 4.0) 2.7 (-7.3 to 12.7)	0 (0 to 65)
Mandelcorn 2004	Self-Assessment score	0–5 (0 = worst, 5 = best)	-2.052 (-8.865 to 4.761)	0.05	-7.7 (-18.9 to 3.6) 4.9 (-6.3 to 16.2)	0 (0 to 85)
	Lower extremity ataxia	Fugl-Meyer: 3-point (0 cannot be performed to 2 can be fully performed)	12.494 (-3.155 to 28.142)	0.025	-6.42 (-35.09 to 22.26) 36.76 (8.09 to 65.43)	35 (0 to 77)
	Truncal ataxia	AMTI forceplate®: NR *Berg Balance Scale® 0–56, with a higher score indicating a better performance*	1.196 (-2.866 to 5.257)	0.690	-0.52 (-8.88 to 7.83) 2.20 (-6.16 to 10.55)	0 (0 to 85)
	Upper extremity ataxia	Purdue Pegboard Test®: pegs inserted into the board with each hand in 30 sec *Minnesota Placing Test®: reach out, grasp, and place blocks in a*	-0.498 (-3.546 to 2.550)	0.382	-3.68 (-10.48 to 3.13) 1.42 (-5.39 to 8.23)	0 (0 to 85)

Author Year	Outcome	Definition / Range of the Scales (severity)	*Main Effect*	*Person-level Heterogeneity of Treatment Effect*		
			Fixed Treatment Effect	P for Person Treatment Interaction*	Treatment Effect Range Lower Range (CI) Upper Range (CI)	I-square % (CI)
		specific order				
McQuay 1994	VAS Pain Intensity	0-100 (0 = no pain, 100 = worst possible pain)	-1.094 (-5.572 to 3.383)	0.004	-8.0 (-18.7 to 2.6) 10.1 (-19.0 to 27.0)	0 (0 to 49)
	VAS Relief Intensity	0-100 (0 = no relief, 100 = complete pain relief)	-3.913 (-11.729 to 3.903)	0.038	-28.4 (-45.9 to -10.8) 5.15 (-12.4 to 22.7)	0 (0 to 49)
Miyazaki 1995	Incidence of angina	Either ST-segment elevation or depression at rest	0.496 (-0.206 to 1.199)	0.125	-16.19 (-6455 to 6422) 17.11 (-6422 to 6456)	0 (0 to 60)
Nathan 2006	Emetic episodes per day	complete response (0 episodes/day), major response (1–2 episodes/day), or failure (>2 episodes/day)	-0.095 (-0.514 to 0.325)	0.001	-16.5 (-4577 to 4523) 2.08 (0.61 to 3.55)	59 (6 to 82)
Parodi 1979	Ischemic attacks	ST elevation or depression (details NR)	-1.544 (-1.838 to -1.251)	0.007	-16.21 (-2668 to 2636) -0.34 (-0.96 to 0.28)	48 (0 to 73)
Parodi 1986	Asymptomatic ST elevation (After verapamil)	0.1 mV of ST-segment elevation measured 20 ms after the J point	-1.637 (-1.994 to -1.279)	0.110	-2.37 (-2.97 to -1.78) -1.30 (-1.74 to -0.86)	6 (0 to 65)
	Asymptomatic ST depression (After verapamil)	More than 0.2 mV of ST-segment depression measured 80 ms after the J point	-1.083 (-1.903 to -0.262)	0.401	-17.42 (-10324 to 10289) -0.90 (-1.81 to 0.00)	0 (0 to 62)
	Symptomatic ST elevation (After verapamil)		-1.580 (-1.906 to -1.254)	<0.001	-15.40 (-3085 to -3054) -1.45 (-1.94 to -0.97)	0 (0 to 62)
	Symptomatic ST Depression (After verapamil)		-0.990 (-1.411 to -0.569)	0.002	-2.53 (-4.25 to -0.80) -0.52 (-2.09 to 1.06)	6 (0 to 64)
	Asymptomatic ST elevation (After propranolol)		0.100 (-0.086 to 0.286)	0.006	-0.77 (-1.72 to 0.18) 1.38 (0.64 to 1.65)	62 (25 to 81)
	Asymptomatic ST depression (After propranolol)		0.339 (-0.168 to 0.845)	0.964	-18.3 (-21040 to 21004) 0.83 (0.01 to 1.64)	0 (0 to 62)
	Symptomatic ST elevation (After propranolol)		-0.002 (-0.177 to 0.173)	0.063	-14.9 (-3159 to 3129) 0.68 (0.34 to 1.02)	46 (0 to 74)
	Symptomatic ST Depression (After propranolol)		-0.374 (-0.709 to -0.039)	0.023	-17.1 (-4612 to 4577) 0.73 (-0.45 to 1.92)	4 (0 to 64)
Pereira 1995	INR	Target INR range of 2.0–3.0	-0.126 (-0.312 to 0.060)	0.433	-0.42 (-1.27 to 0.08) 0.16 (-0.28 to 0.59)	0 (0 to 71)
Tison 2012	Troublesome dyskinesia	7 points scale (1 = extremely uncomfortable, 7 = not at all uncomfortable)	0.167 (-0.449 to 0.783)	0.593	-0.67 (-2.68 to 1.35) 1.83 (-0.18 to 3.85)	0 (0 to 62)

* The significance of person-level HTE was assessed by a likelihood ratio test comparing the two models – model with common treatment effect and model with treatment-by-participant interactions

Figure Legend

Figure 1: The Figure provides a schematic description of: person-level outcomes (outcomes for each patient during each treatment period); person-level effects (contrasts of the outcomes for each patient in one treatment condition *versus* another); and person-HTE (between patient contrasts of effects).

Figure 2. Study Flow Diagram represents the flow of eligible studies included in this review

Figure 3. Person-level variation across different disease conditions. This figure depicts the results of 46 different N-of-1 trials of cimetidine as reported by Johanessen et al [9]. The effect of cimetidine versus placebo was measured in each subject across 12 cross-over periods over the span of 184 days. While cimetidine had a similar average effect regardless of the index condition, there was far greater consistency of effect in patients with peptic ulcer disease and much more variation in effect among patients with non-ulcer dyspepsia.

Figure 1. Person-level outcomes, person-level effects and person-level HTE

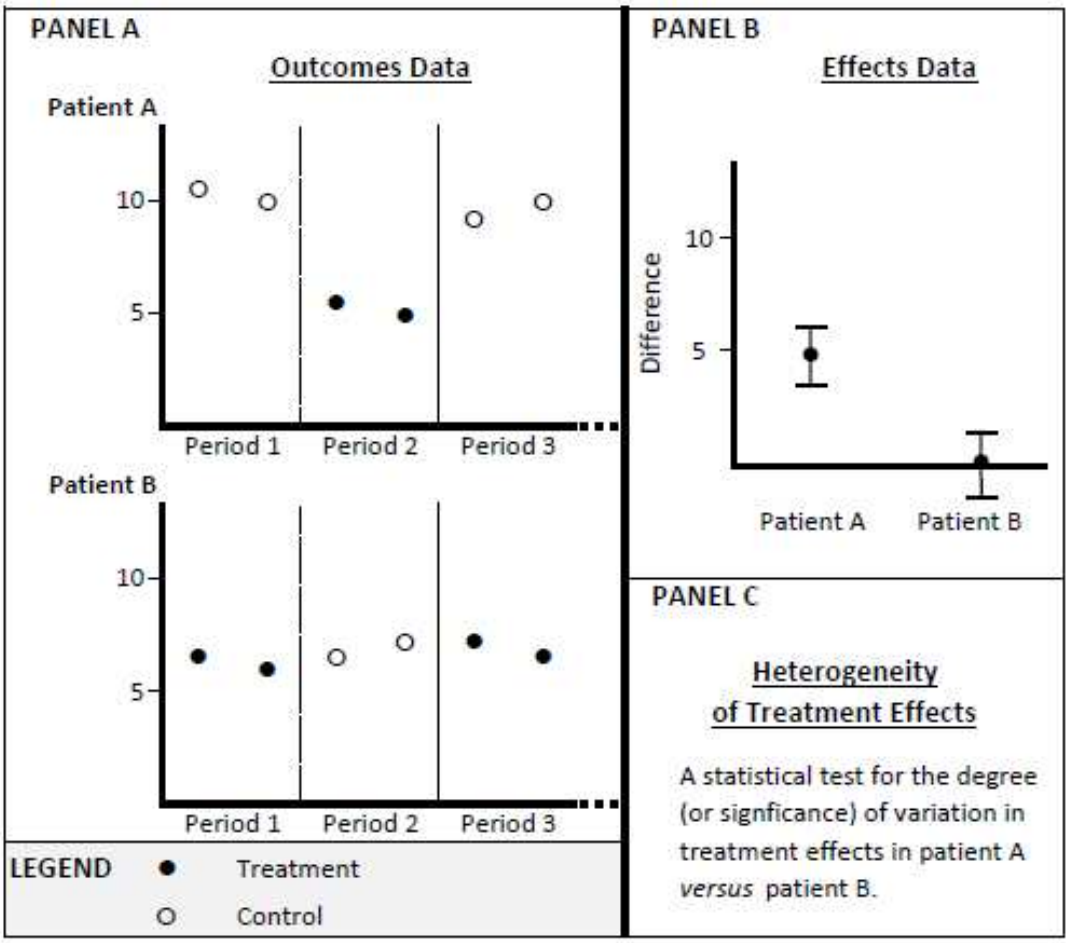


Figure 2. Study Flow Diagram

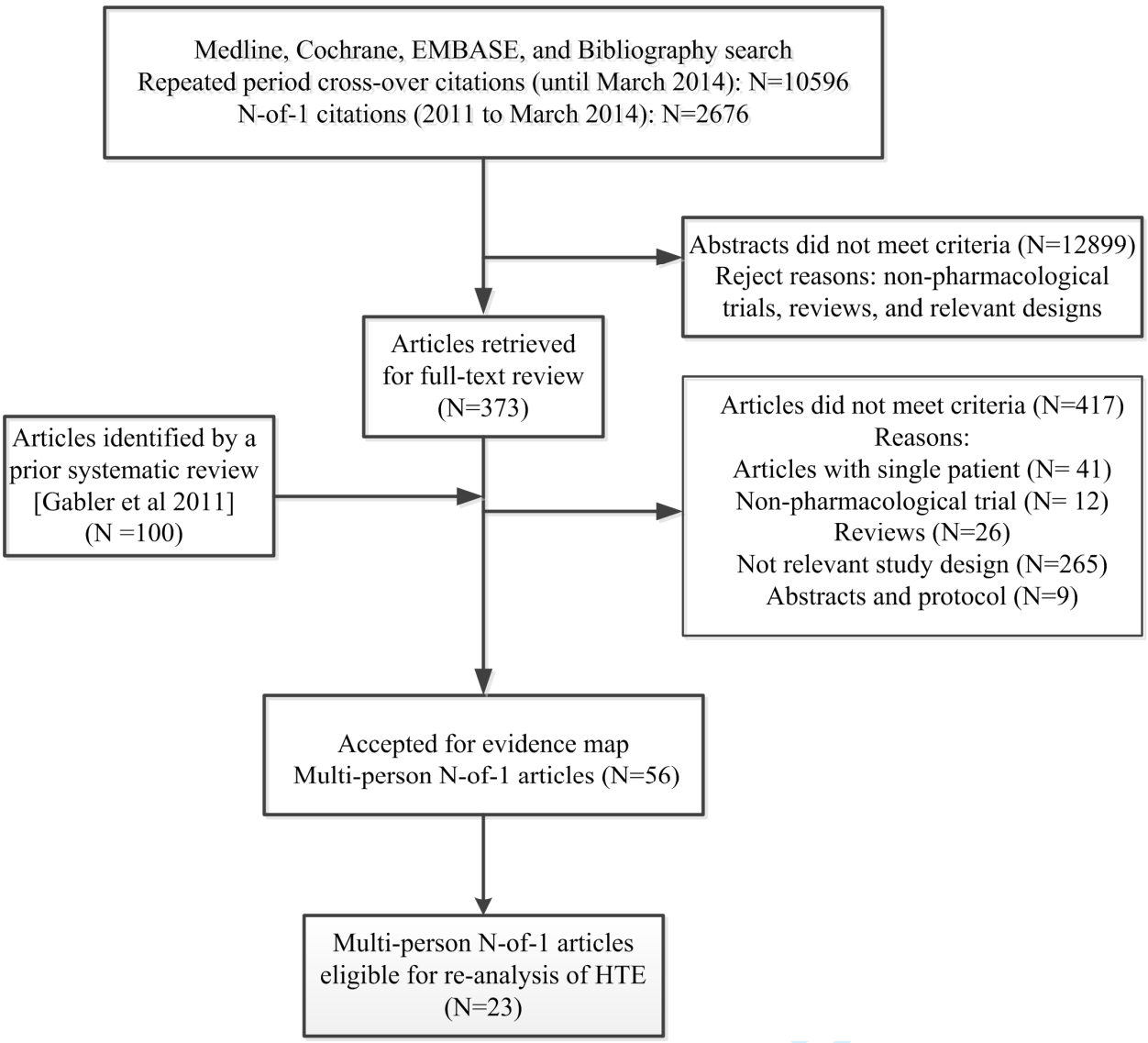
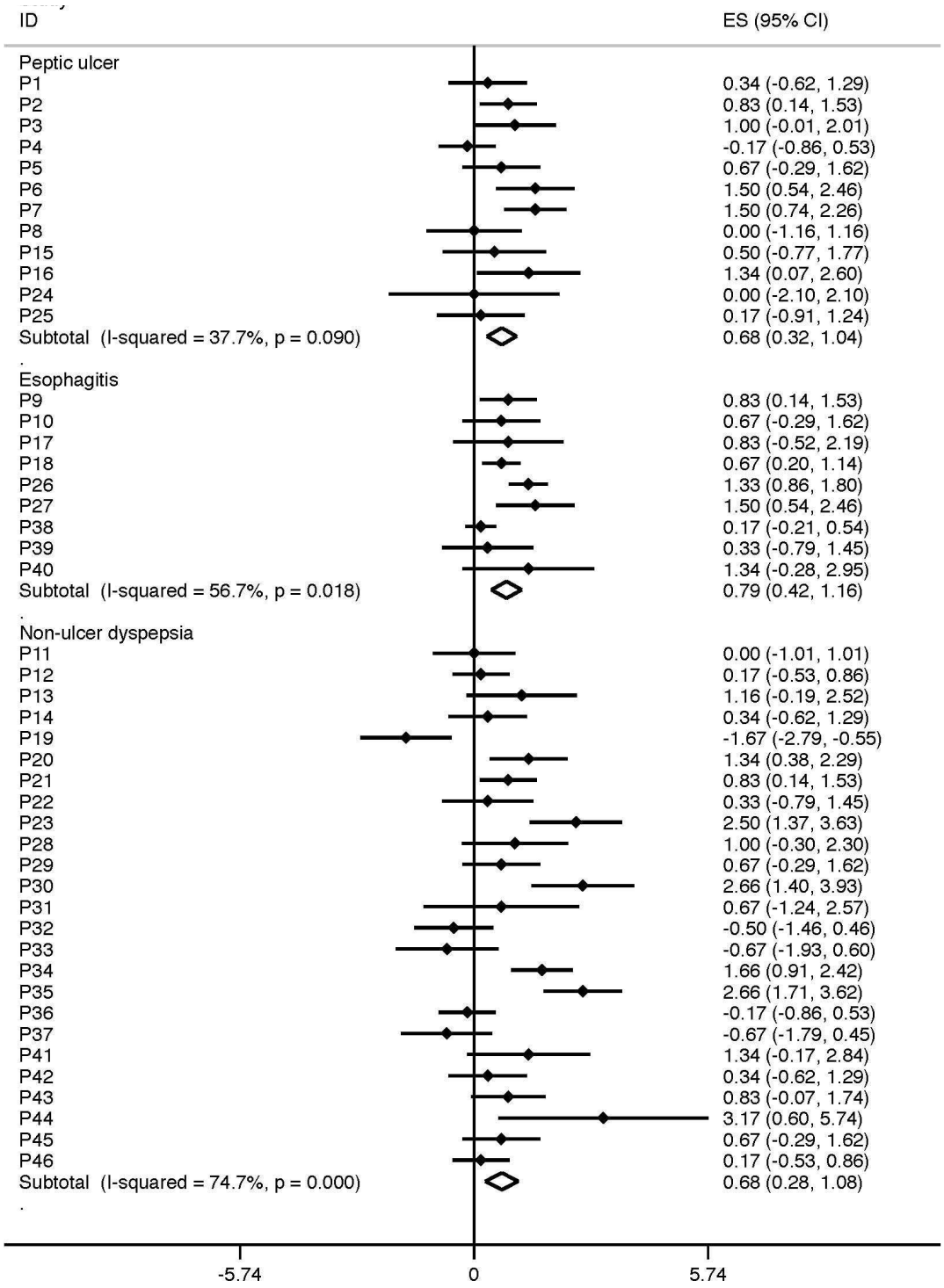


Figure 3. Person-level variation across different disease conditions



Appendix Materials

Appendix Table 1: N-of-1 Trial Searches

1.	randomized controlled trial.pt.
2.	controlled clinical trial.pt.
3.	randomized controlled trials/
4.	Double-blind Method/
5.	Single-Blind Method/
6.	clinical trial.pt.
7.	Clinical Trials.mp. or exp Clinical Trials/
8.	random$.tw.
9.	trial$.tw.
10.	Cross-Over Studies/
11.	or/1-10
12.	n-of-1.af.
13.	11 and 12
14.	(single-subject or single-patient or single case or single-case or within-patient).af.
15.	((single adj1 patient) or (single adj1 subject)).tw.
16.	14 or 15
17.	11 and 16
18.	multi-crossover.mp.
19.	13 or 17 or 18
20.	limit 19 to yr="2010 - 2014"

Appendix Table 2: Repeated Period Crossover Trials

1.	(repeat$ or rotat$).af.
2.	((three or four or five or six) and period).tw.
3.	(multi- or multiple).tw.
4.	(three-period or four-period or five-period or six-period).tw.
5.	(three-way or four-way or five-way or six-way).tw.
6.	or/1-5
7.	Cross-Over Studies/ or (cross-over or crossover).af.
8.	6 and 7
9.	randomized controlled trial.pt.
10.	controlled clinical trial.pt.
11.	randomized controlled trials/
12.	Double-blind Method/
13.	Single-Blind Method/
14.	clinical trial.pt.
15.	Clinical Trials.mp. or exp Clinical Trials/
16.	random$.tw.
17.	trial$.tw.
18.	or/9-17
19.	8 and 18
20.	(dt or de or tu).fs.
21.	19 and 20
22.	7 and 20
23.	"Reproducibility of Results"/
24.	16 and 22
25.	limit 22 to english language
26.	9 or 10 or 11 or 14 or 15 or 16
27.	7 or 23
28.	20 and 26 and 27
29.	random.af.
30.	9 or 10 or 11 or 14 or 15 or 29
31.	ae.fs.
32.	20 or 31
33.	27 and 30 and 32
34.	limit 33 to (english language and humans)
35.	periods.af.
36.	6 or 35

37.	33 and 36
38.	Animals/ not human/
39.	37 not 38

Appendix Table 3: Reference List (Included N-of-1 Studies)

1.	Nikles CJ, McKinlay L, Mitchell GK, Carmont SA, Senior HE, Waugh MC et al. Aggregated n-of-1 trials of central nervous system stimulants versus placebo for paediatric traumatic brain injury--a pilot study. Trials [Electronic Resource] 2014; 15:54.
2.	Tison F, Negre-Pages L, Meissner WG, Dupouy S, Li Q, Thiolat ML et al. Simvastatin decreases levodopa-induced dyskinesia in monkeys, but not in a randomized, placebo-controlled, multiple cross-over ("n-of-1") exploratory trial of simvastatin against levodopa-induced dyskinesia in Parkinson's disease patients. Parkinsonism & Related Disorders 2013; 19(4):416-421.
3.	Rascol O, Ferreira J, Negre-Pages L, Perez-Lloret S, Lacomblez L, Galitzky M et al. A proof-of-concept, randomized, placebo-controlled, multiple cross-overs (n-of-1) study of naftazone in Parkinson's disease. Fundamental & Clinical Pharmacology 2012; 26(4):557-564.
4.	Emmanuel AV, Kamm MA, Roy AJ, Kerstens R, Vandeplassche L. Randomised clinical trial: the efficacy of prucalopride in patients with chronic intestinal pseudo-obstruction--a double-blind, placebo-controlled, cross-over, multiple n = 1 study. Alimentary Pharmacology & Therapeutics 2012; 35(1):48-55.
5.	Yelland MJ, Poulos CJ, Pillans PI, Bashford GM, Nikles CJ, Sturtevant JM et al. N-of-1 randomized trials to assess the efficacy of gabapentin for chronic neuropathic pain. Pain Medicine 2009; 10(4):754-761.
6.	Nonoyama ML, Brooks D, Guyatt GH, Goldstein RS. Effect of oxygen on health quality of life in patients with chronic obstructive pulmonary disease with transient exertional hypoxemia. American Journal of Respiratory & Critical Care Medicine 2007; 176(4):343-349.
7.	Huber AM, Tomlinson GA, Koren G, Feldman BM. Amitriptyline to relieve pain in juvenile idiopathic arthritis: a pilot study using Bayesian metaanalysis of multiple N-of-1 clinical trials. Journal of Rheumatology 2007; 34(5):1125-1132.
8.	Yelland MJ, Nikles CJ, McNairn N, Del Mar CB, Schluter PJ, Brown RM. Celecoxib compared with sustained-release paracetamol for osteoarthritis: a series of n-of-1 trials. Rheumatology 2007; 46(1):135-140.
9.	Zucker DR, Ruthazer R, Schmid CH, Feuer JM, Fischer PA, Kieval RI et al. Lessons learned combining N-of-1 trials to assess fibromyalgia therapies. Journal of Rheumatology 2006; 33(10):2069-2077.
10.	Nikles CJ, Mitchell GK, Del Mar CB, Clavarino A, McNairn N. An n-of-1 trial service in clinical practice: testing the effectiveness of stimulants for attention-deficit/hyperactivity disorder. Pediatrics 2006; 117(6):2040-2046.
11.	Nathan PC, Tomlinson G, Dupuis LL, Greenberg ML, Ota S, Bartels U et al. A pilot study of ondansetron plus metopimazine vs. ondansetron monotherapy in children receiving highly emetogenic chemotherapy: a Bayesian randomized serial N-of-1 trials design. Supportive Care in Cancer 2006; 14(3):268-276.
12.	Pereira JA, Holbrook AM, Dolovich L, Goldsmith C, Thabane L, Douketis JD et al. Are brand-name and generic warfarin interchangeable? Multiple n-of-1 randomized, crossover trials. Annals of Pharmacotherapy 2005; 39(7-8):1188-1193.
13.	Woodfield R, Goodyear-Smith F, Arroll B. N-of-1 trials of quinine efficacy in skeletal muscle cramps of the leg. British Journal of General Practice 2005; 55(512):181-185.
14.	Wegman AC, van der Windt DA, Bongers M, Twisk JW, Stalman WA, de Vries TP. Efficacy of temazepam in frequent users: a series of N-of-1 trials. Family Practice 2005; 22(2):152-159.
15.	Nikles CJ, Yelland M, Glasziou PP, Del MC. Do individualized medication effectiveness tests (n-of-1 trials) change clinical decisions about which drugs to use for osteoarthritis and chronic pain?. [Review] [19 refs]. American Journal of Therapeutics 2005; 12(1):92-97.
16.	Smith BJ, Appleton SL, Veale AJ, McElroy HJ, Veljkovic D, Saccoia L. Eformoterol n-of-1 trials in chronic obstructive pulmonary disease poorly reversible to salbutamol. Chronic Respiratory Disease 2004; 1(2):63-69.

17.	Haas DC, Sheehe PR. Dextroamphetamine pilot crossover trials and n of 1 trials in patients with chronic tension-type and migraine headache. Headache 2004; 44(10):1029-1037.
18.	Mandelcorn J, Cullen NK, Bayley MT. A preliminary study of the efficacy of ondansetron in the treatment of ataxia, poor balance and incoordination from brain injury. Brain Injury 2004; 18(10):1025-1039.
19.	Pope JE, Prashker M, Anderson J. The efficacy and cost effectiveness of N of 1 studies with diclofenac compared to standard treatment with nonsteroidal antiinflammatory drugs in osteoarthritis. Journal of Rheumatology 2004; 31(1):140-149.
20.	Wegman AC, van der Windt DA, de HM, Deville WL, Fo CT, de Vries TP. Switching from NSAIDs to paracetamol: a series of n of 1 trials for individual patients with osteoarthritis. Annals of the Rheumatic Diseases 2003; 62(12):1156-1161.
21.	Reitberg DP, Del RE, Weiss SL, Rebell G, Zaias N. Single-patient drug trial methodology for allergic rhinitis. Annals of Pharmacotherapy 2002; 36(9):1366-1374.
22.	Linday LA, Tsiouris JA, Cohen IL, Shindledecker R, DeCresce R. Famotidine treatment of children with autistic spectrum disorders: pilot research using single subject research design. Journal of Neural Transmission 2001; 108(5):593-611.
23.	Duggan CM, Mitchell G, Nikles CJ, Glasziou PP, Del Mar CB, Clavarino A. Managing ADHD in general practice. N of 1 trials can help! Australian Family Physician 2000; 29(12):1205-1209.
24.	Nikles CJ, Glasziou PP, Del Mar CB, Duggan CM, Clavarino A, Yelland MJ. Preliminary experiences with a single-patient trials service in general practice. Medical Journal of Australia 2000; 173(2):100-103.
25.	Mahon JL, Laupacis A, Hodder RV, McKim DA, Paterson NA, Wood TE et al. Theophylline for irreversible chronic airflow limitation: a randomized study comparing n of 1 trials to standard practice. Chest 1999; 115(1):38-48.
26.	Bollert FG, Paton JY, Marshall TG, Calvert J, Greening AP, Innes JA. Recombinant DNase in cystic fibrosis: a protocol for targeted introduction through n-of-1 trials. Scottish Cystic Fibrosis Group. European Respiratory Journal 1999; 13(1):107-113.
27.	Kent MA, Camfield CS, Camfield PR. Double-blind methylphenidate trials: practical, useful, and highly endorsed by families. Archives of Pediatrics & Adolescent Medicine 1999; 153(12):1292-1296.
28.	Webb S, Tansey P, Brown H, Jackson A, Bilton D. Placebo-controlled n-of-1 trials in cystic fibrosis. European Respiratory Journal 1999; 14(4):993.
29.	Haines DR, Gaines SP. N of 1 randomised controlled trials of oral ketamine in patients with chronic pain. Pain 1999; 83(2):283-287.
30.	Sheather-Reid RB, Cohen M. Efficacy of analgesics in chronic pain: a series of N-of-1 studies. Journal of Pain & Symptom Management 1998; 15(4):244-252.
31.	Camfield P, Gordon K, Dooley J, Camfield C. Melatonin appears ineffective in children with intellectual deficits and fragmented sleep: six "N of 1" trials. Journal of Child Neurology 1996; 11(4):341-343.
32.	Mahon J, Laupacis A, Donner A, Wood T. Randomised study of n of 1 trials versus standard practice.[Erratum appears in BMJ 1996 Jun 1;312(7043):1392]. BMJ 1996; 312(7038):1069-1074.
33.	Maier W, Benkert O. Treatment of chronic depression with sulpiride: evidence of efficacy in placebo-controlled single case studies. Psychopharmacology 1994; 115(4):495-501.
34.	McQuay HJ, Carroll D, Jadad AR, Glynn CJ, Jack T, Moore RA et al. Dextromethorphan for the treatment of neuropathic pain: a double-blind randomised controlled crossover trial with integral n-of-1 design. Pain 1994; 59(1):127-133.
35.	March L, Irwig L, Schwarz J, Simpson J, Chock C, Brooks P. n of 1 trials comparing a non-steroidal anti-inflammatory drug with paracetamol in osteoarthritis. BMJ 1045; 309(6961):1041-1045.
36.	Denburg SD, Carbotte RM, Denburg JA. Corticosteroids and neuropsychological functioning in

	patients with systemic lupus erythematosus. Arthritis & Rheumatism 1994; 37(9):1311-1320.
37.	Privitera MD, Treiman DM, Pledger GW, Sahlroot JT, Handforth A, Linde MS et al. Dezinamide for partial seizures: results of an n-of-1 design trial. Neurology 1994; 44(8):1453-1458.
38.	Langer JC, Winthrop AL, Issenman RM. The single-subject randomized trial. A useful clinical tool for assessing therapeutic efficacy in pediatric practice. Clinical Pediatrics 1993; 32(11):654-657.
39.	Molloy DW, Guyatt GH, Standish T, Willan A, McIlroy W, D'Souza J et al. Effect of a new nootropic agent, CGS 5649B, on cognition, function, and behavior in dementia. Journal of General Internal Medicine 1993; 8(8):444-447.
40.	Johannessen T, Petersen H, Kristensen P, Fosstvedt D, Kleveland PM, Dybdahl J et al. Cimetidine on-demand in dyspepsia. Experience with randomized controlled single-subject trials. Scandinavian Journal of Gastroenterology 1992; 27(3):189-195.
41.	Johannessen T, Kristensen P, Petersen H, Fosstvedt D, Loge I, Kleveland PM et al. The symptomatic effect of 1-day treatment periods with cimetidine in dyspepsia. Combined results from randomized, controlled, single-subject trials. Scandinavian Journal of Gastroenterology 1991; 26(9):974-980.
42.	Patel A, Jaeschke R, Guyatt GH, Keller JL, Newhouse MT. Clinical usefulness of n-of-1 randomized controlled trials in patients with nonreversible chronic airflow limitation. American Review of Respiratory Disease 1991; 144(4):962-964.
43.	Larsen S, Farup P, Flaten O, Osnes M. The multi-crossover model for classifying patients as responders to a given treatment. Scandinavian Journal of Gastroenterology 1991; 26(7):763-770.
44.	Jaeschke R, Adachi J, Guyatt G, Keller J, Wong B. Clinical usefulness of amitriptyline in fibromyalgia: the results of 23 N-of-1 randomized controlled trials. Journal of Rheumatology 1991; 18(3):447-451.
45.	Hinderer SR. The supraspinal anxiolytic effect of baclofen for spasticity reduction. American Journal of Physical Medicine & Rehabilitation 1990; 69(5):254-258.
46.	Lashner BA, Hanauer SB, Silverstein MD. Testing nicotine gum for ulcerative colitis patients. Experience with single-patient trials. Digestive Diseases & Sciences 1990; 35(7):827-832.
47.	McBride MC. An individual double-blind crossover trial for assessing methylphenidate response in children with attention deficit disorder. Journal of Pediatrics 1988; 113(1:Pt 1):t-45.
48.	Menard J, Serrurier D, Bautier P, Plouin PF, Corvol P. Crossover design to test antihypertensive drugs with self-recorded blood pressure. Hypertension 1988; 11(2):153-159.
49.	Ullmann RK, Sleator EK. Responders, nonresponders, and placebo responders among children with attention deficit disorder. Importance of a blinded placebo evaluation. Clinical Pediatrics 1986; 25(12):594-599.
50.	Wolfe B, Del RE, Weiss SL, Mendelson A, Elbaga TA, Huser FJ et al. Validation of a single-patient drug trial methodology for personalized management of gastroesophageal reflux disease. J Manag Care Pharm 2002; 8(6):459-468.
51.	Brookes ST, Biddle L, Paterson C, Woolhead G, Dieppe P. "Me's me and you's you": Exploring patients' perspectives of single patient (n-of-1) trials in the UK. Trials 2007; 8:10.
52.	Wallace AE, Kofoed LL. Statistical Analysis of Single Case Studies in the Clinical Setting: The Example of Methylphenidate Trials in Children with Attention-Deficit Hyperactivity Disorder. Journal of Child and Adolescent Psychopharmacology 1994; 4(3):141-150.

Appendix Table 4: Reference List (Included Repeated Period Crossover Studies)

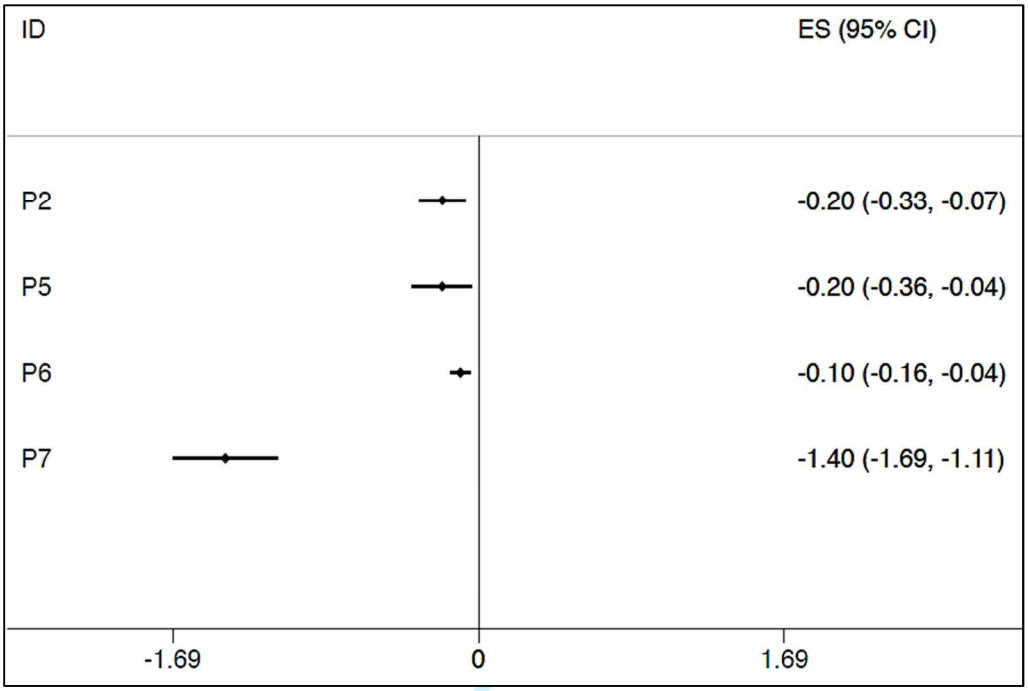
1.	Seeburger JL, Cady RK, Winner P, MacGregor A, Valade D, Ge Y et al. Rizatriptan for treatment of acute migraine in patients taking topiramate for migraine prophylaxis. Headache 2012; 52(1):57-67.
2.	Fallon M, Reale C, Davies A, Lux AE, Kumar K, Stachowiak A et al. Efficacy and safety of fentanyl pectin nasal spray compared with immediate-release morphine sulfate tablets in the treatment of breakthrough cancer pain: a multicenter, randomized, controlled, double-blind, double-dummy multiple-crossover study. The Journal of Supportive Oncology 2011; 9(6):224-231.
3.	Allais G, Bussone G, D'Andrea G, Moschiano F, d'Onofrio F, Valguarnera F et al. Almotriptan 12.5 mg in menstrually related migraine: a randomized, double-blind, placebo-controlled study. Cephalalgia 2011; 31(2):144-151.
4.	Farrar JT, Messina J, Xie F, Portenoy RK. A novel 12-week study, with three randomized, double-blind placebo-controlled periods to evaluate fentanyl buccal tablets for the relief of breakthrough pain in opioid-tolerant patients with noncancer-related chronic pain. Pain Medicine 2010; 11(9):1313-1327.
5.	Stocchi F, Hersh BP, Scott BL, Nausieda PA, Giorgi L, Ease-PD Monotherapy Study Investigators. Ropinirole 24-hour prolonged release and ropinirole immediate release in early Parkinson's disease: a randomized, double-blind, non-inferiority crossover study. Current Medical Research & Opinion 2008; 24(10):2883-2895.
6.	Incrocci L, Slob AK, Hop WC. Tadalafil (Cialis) and erectile dysfunction after radiotherapy for prostate cancer: an open-label extension of a blinded trial. Urology 2007; 70(6):1190-1193.
7.	Danielsson P, Janson A, Norgren S, Marcus C. Impact sibutramine therapy in children with hypothalamic obesity or obesity with aggravating syndromes. Journal of Clinical Endocrinology & Metabolism 2007; 92(11):4101-4106.
8.	Lu M, Malladi V, Agha A, Abudayyeh S, Han C, Siepman N et al. Failures in a proton pump inhibitor therapeutic substitution program: lessons learned. Digestive Diseases & Sciences 2007; 52(10):2813-2820.
9.	Horiguchi T, Hayashi N, Ohira D, Torigoe H, Ito T, Hirose M et al. Usefulness of HFA-BDP for adult patients with bronchial asthma: randomized crossover study with fluticasone. Journal of Asthma 2006; 43(7):509-512.
10.	Lauschke A, Teichgraber UK, Frei U, Eckardt KU. 'Low-dose' dopamine worsens renal perfusion in patients with acute renal failure. Kidney International 2006; 69(9):1669-1674.
11.	Carpentier PJ, de Jong CA, Dijkstra BA, Verbrugge CA, Krabbe PF. A controlled trial of methylphenidate in adults with attention deficit/hyperactivity disorder and substance use disorders. Addiction 2005; 100(12):1868-1874.
12.	Bergemann N, Mundt C, Parzer P, Pakrasi M, Eckstein-Mannsperger U, Haisch S et al. Estrogen as an adjuvant therapy to antipsychotics does not prevent relapse in women suffering from schizophrenia: results of a placebo-controlled double-blind study. Schizophrenia Research 2005; 74(2-3):125-134.
13.	Rao U, Ott GE, Lin KM, Gertsik L, Poland RE. Effect of bupropion on nocturnal urinary free cortisol and its association with antidepressant response. Journal of Psychiatric Research 2005; 39(2):183-190.
14.	Whyte J, Hart T, Vaccaro M, Grieb-Neff P, Risser A, Polansky M et al. Effects of methylphenidate on attention deficits after traumatic brain injury: a multidimensional, randomized, controlled trial. American Journal of Physical Medicine & Rehabilitation 2004; 83(6):401-420.
15.	Incrocci L, Hop WC, Slob AK. Efficacy of sildenafil in an open-label study as a continuation of a double-blind study in the treatment of erectile dysfunction after radiotherapy for prostate cancer. Urology 2003; 62(1):116-120.
16.	Deary AJ, Schumann AL, Murfet H, Haydock S, Foo RS, Brown MJ. Influence of drugs and gender on the arterial pulse wave and natriuretic peptide secretion in untreated patients with

	essential hypertension. Clinical Science 2002; 103(5):493-499.
17.	Reitberg DP, Del RE, Weiss SL, Rebell G, Zaias N. Single-patient drug trial methodology for allergic rhinitis. Annals of Pharmacotherapy 2002; 36(9):1366-1374.
18.	White JM, Danz C, Kneebone J, La Vincente SF, Newcombe DA, Ali RL. Relationship between LAAM-methadone preference and treatment outcomes. Drug & Alcohol Dependence 2002; 66(3):295-301.
19.	Deary AJ, Schumann AL, Murfet H, Haydock SF, Foo RS, Brown MJ. Double-blind, placebo-controlled crossover comparison of five classes of antihypertensive drugs. Journal of Hypertension 2002; 20(4):771-777.
20.	Juan O, Campos JM, Caranana V, Sanchez JJ, Casan R, Alberola V. A randomized, crossover comparison of standard-dose versus low-dose lenograstim in the prophylaxis of post-chemotherapy neutropenia. Supportive Care in Cancer 2001; 9(4):241-246.
21.	Koran LM, Gelenberg AJ, Kornstein SG, Howland RH, Friedman RA, DeBattista C et al. Sertraline versus imipramine to prevent relapse in chronic depression. Journal of Affective Disorders 2001; 65(1):27-36.
22.	Dasbach EJ, Carides GW, Gerth WC, Santanello NC, Pigeon JG, Kramer. Work and productivity loss in the rizatriptan multiple attack study. Cephalalgia 2000; 20(9):830-834.
23.	Dickerson JE, Hingorani AD, Ashby MJ, Palmer CR, Brown MJ. Optimisation of antihypertensive treatment by crossover rotation of four major classes. Lancet 1999; 353(9169):2008-2013.
24.	Martinek J, Blum AL, Stolte M, Hartmann M, Verdu EF, Luhmann R et al. Effects of pumaprazole (BY841), a novel reversible proton pump antagonist, and of omeprazole, on intragastric acidity before and after cure of Helicobacter pylori infection. Alimentary Pharmacology & Therapeutics 1999; 13(1):27-34.
25.	Weiss R, Ferry D, Pickering E, Smith LK, Dennish G, III, Krug-Gourley S et al. Effectiveness of three different doses of carvedilol for exertional angina. Carvedilol-Angina Study Group. American Journal of Cardiology 1998; 82(8):927-931.
26.	Whyte J, Hart T, Schuster K, Fleming M, Polansky M, Coslett HB. Effects of methylphenidate on attentional function after traumatic brain injury. A randomized, placebo-controlled trial. American Journal of Physical Medicine & Rehabilitation 1997; 76(6):440-450.
27.	Walker JS, Sheather-Reid RB, Carmody JJ, Vial JH, Day RO. Nonsteroidal antiinflammatory drugs in rheumatoid arthritis and osteoarthritis: support for the concept of "responders" and "nonresponders". Arthritis & Rheumatism 1997; 40(11):1944-1954.
28.	Ernst DS, Brasher P, Hagen N, Paterson AH, MacDonald RN, Bruera E. A randomized, controlled trial of intravenous clodronate in patients with metastatic bone disease and pain. Journal of Pain & Symptom Management 1997; 13(6):319-326.
29.	O'Driscoll G, Green D, Taylor RR. Simvastatin, an HMG-coenzyme A reductase inhibitor, improves endothelial function within 1 month. Circulation 1997; 95(5):1126-1131.
30.	Collins JJ, Geake J, Grier HE, Houck CS, Thaler HT, Weinstein HJ et al. Patient-controlled analgesia for mucositis pain in children: a three-period crossover study comparing morphine and hydromorphone. Journal of Pediatrics 1996; 129(5):722-728.
31.	Boolell M, Gepi-Attee S, Gingell JC, Allen MJ. Sildenafil, a novel effective oral therapy for male erectile dysfunction. British Journal of Urology 1996; 78(2):257-261.
32.	Wysocki M, Andersson OK, Persson B, Bagge U, Braide M. Hemorheologic effects of vasodilation in essential hypertension. Angiology 1996; 47(9):869-878.
33.	Miyazaki S, Nonogi H, Goto Y, Sumiyoshi T, Haze K, Hiramori K. Comparison of the therapeutic efficacy of continuous and intermittent injection of isosorbide dinitrate: a randomized study on unstable angina. Internal Medicine 1995; 34(9):856-862.
34.	Meeves S, Hafner K, Park G, Weber M. Three-period crossover trial with ambulatory blood pressure monitoring for evaluating antihypertensive therapy. American Heart Journal 1995; 130(4):841-848.

35.	Rederich G, Rapoport A, Cutler N, Hazelrigg R, Jamerson B. Oral sumatriptan for the long-term treatment of migraine: clinical findings. Neurology 1995; 45(8:Suppl 7):Suppl-20.
36.	Stenberg A, Lackgren G. Desmopressin tablets in the treatment of severe nocturnal enuresis in adolescents. Pediatrics 1994; 94(6:Pt 1):t-6.
37.	Van den Eeden SK, Koepsell TD, Longstreth WT, Jr., van BG, Daling JR, McKnight B. Aspartame ingestion and headaches: a randomized crossover trial. Neurology 1994; 44(10):1787-1793.
38.	Langeland T, Fagertun HE, Larsen S. Therapeutic effect of loratadine on pruritus in patients with atopic dermatitis. A multi-crossover-designed study. Allergy 1994; 49(1):22-26.
39.	Campbell SH, Hickey-Dwyer M, Harding SP. Double-masked three-period crossover investigation of timolol in control of raised intraocular pressure. Eye 1993; 7(Pt 1):105-108.
40.	Johnsen V, Brun JG, Fjeld E, Hansen K, Sydnes OA, Ugstad MB. Morning stiffness and nightime pain in ankylosing spondylitis. A comparison between enteric-coated and plain naproxen tablets. European Journal of Rheumatology & Inflammation 1992; 12(2):37-42.
41.	Terho P. Desmopressin in nocturnal enuresis. Journal of Urology 1991; 145(4):818-820.
42.	Molloy DW, Guyatt GH, Wilson DB, Duke R, Rees L, Singer J. Effect of tetrahydroaminoacridine on cognition, function and behaviour in Alzheimer's disease. CMAJ Canadian Medical Association Journal 1991; 144(1):29-34.
43.	Greco R, Schiattarella M, Wolff S, D'Alterio D, Tartaglia P. Long-term efficacy of nitroglycerin patch in stable angina pectoris. American Journal of Cardiology 1990; 65(21):9J-15J.
44.	Sabesin SM, Boyce HW, Jr., King CE, Mann JA, Ruoff G, Wall E. Comparative evaluation of gastrointestinal intolerance produced by plain and tri-buffered aspirin tablets. American Journal of Gastroenterology 1988; 83(11):1220-1225.
45.	Johannessen T, Fjosne U, Kleveland PM, Halvorsen T, Kristensen P, Loge I et al. Cimetidine responders in non-ulcer dyspepsia. Scandinavian Journal of Gastroenterology 1988; 23(3):327-336.
46.	Kauppila A, Makila UM, Makarainen L, Puolakka J, Seppala A. Tiaprofenic acid in the treatment of primary dysmenorrhoea. European Journal of Obstetrics, Gynecology, & Reproductive Biology 1986; 22(5-6):359-363.
47.	Koopmans PP, Thien T, Thomas CM, Van den Berg RJ, Gribnau FW. The effects of sulindac and indomethacin on the anti-hypertensive and diuretic action of hydrochlorothiazide in patients with mild to moderate essential hypertension. British Journal of Clinical Pharmacology 1986; 21(4):417-423.
48.	Parodi O, Simonetti I, Michelassi C, Carpeggiani C, Biagini A, L'Abbate A et al. Comparison of verapamil and propranolol therapy for angina pectoris at rest: a randomized, multiple-crossover, controlled trial in the coronary care unit. American Journal of Cardiology 1986; 57(11):899-906.
49.	Gheorghiade M, St CC, St CJ, Freedman D, Schwemer G. Short- and long-term treatment of stable effort angina with nicardipine, a new calcium channel blocker: a double-blind, placebo-controlled, randomised, repeated cross-over study. British Journal of Clinical Pharmacology 1985; 20:Suppl-205S.
50.	Dickstein K, Knutsen H. A double-blind multiple crossover trial evaluating a transdermal nitroglycerin system vs placebo. European Heart Journal 1985; 6(1):50-56.
51.	Kleveland PM, Larsen S, Sandvik L, Kristensen P, Johannessen T, Hafstad PE et al. The effect of cimetidine in non-ulcer dyspepsia. Experience with a multi-cross-over model. Scandinavian Journal of Gastroenterology 1985; 20(1):19-24.
52.	Kauppila A, Ronnberg L. Naproxen sodium in dysmenorrhea secondary to endometriosis. Obstetrics & Gynecology 1985; 65(3):379-383.
53.	Deroover J, Baro F, Bourguignon RP, Smets P. Tiapride versus placebo: a double-blind comparative study in the management of Huntington's chorea. Current Medical Research & Opinion 1984; 9(5):329-338.
54.	Crean PA, Ribeiro P, Crea F, Davies GJ, Ratcliffe D, Maseri A. Failure of transdermal nitroglycerin to improve chronic stable angina: a randomized, placebo-controlled, double-blind,

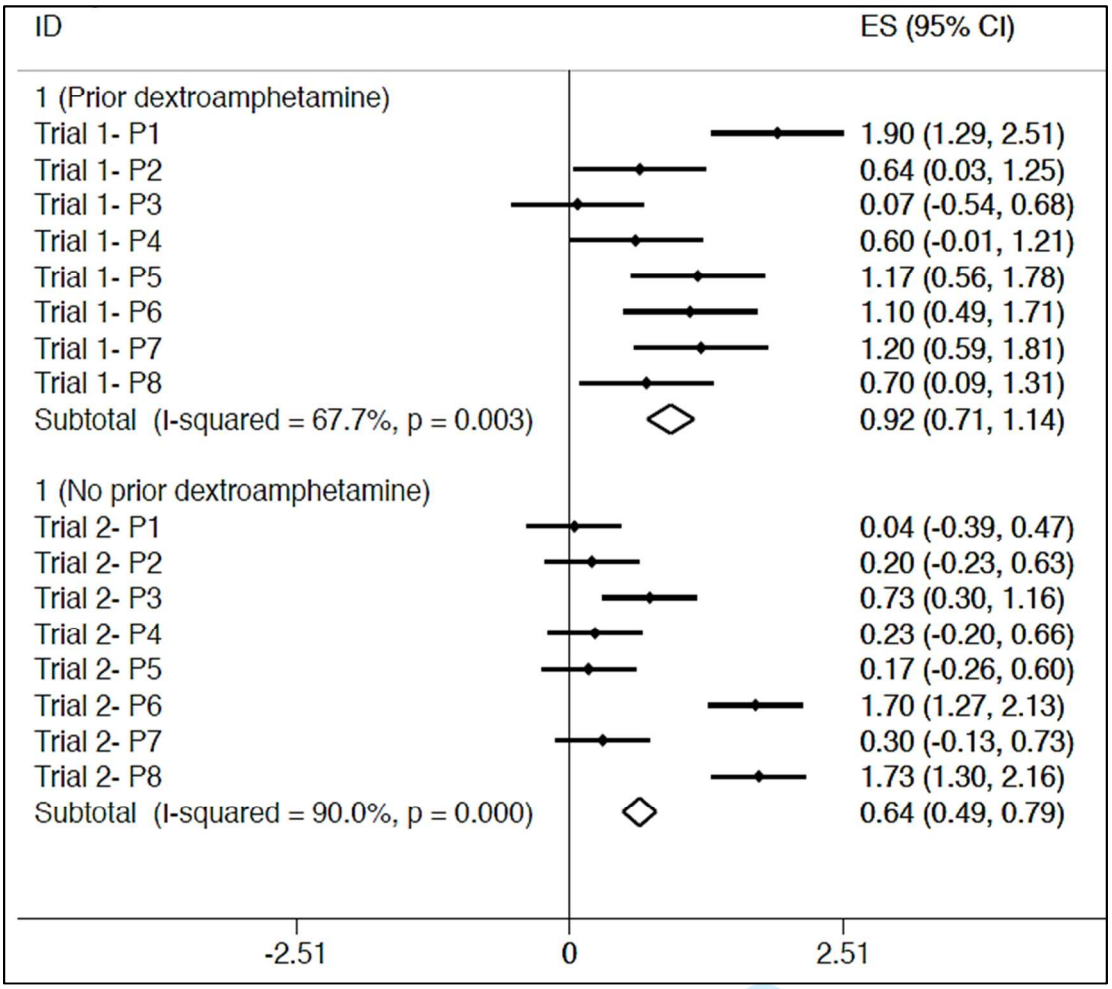
	double crossover trial. American Heart Journal 1984; 108(6):1494-1500.
55.	De CR, Carpeggiani C, L'Abbate A. A double-blind, placebo-controlled study of ketanserin in patients with Prinzmetal's angina. Evidence against a role for serotonin in the genesis of coronary vasospasm. Circulation 1984; 69(5):889-894.
56.	Hunter JR, Galloway JR, Brooke MM, Kutner MH, Rudman D, Vogel RL et al. Effects of allopurinol in Duchenne's muscular dystrophy. Archives of Neurology 1983; 40(5):294-299.
57.	Boissl K, Dreyfus JF, Delmotte M. Studies on the dependence-inducing potential of zopiclone and triazolam. International Pharmacopsychiatry 1982; 17:Suppl-7.
58.	Glass RM, Uhlenhuth EH, Matuzas W, McCracken S, Greene S. Subject-own-control design in evaluating clinical antidepressant effects. Journal of Affective Disorders 1982; 4(4):373-381.
59.	Damsgaard EM, Faber O, Froland A, Iversen S. Therapeutic effect of tolbutamide in non-insulin dependent diabetes mellitus (NIDDM). Relation to beta-cell function. Acta Endocrinologica 1982; 100(3):416-420.
60.	Riihiluoma P, Wuolijoki E, Pulkkinen MO. Treatment of primary dysmenorrhea with diclofenac sodium. European Journal of Obstetrics, Gynecology, & Reproductive Biology 1981; 12(3):189-194.
61.	Subramanian B, Bowles M, Lahiri A, Davies AB, Raftery EB. Long-term antianginal action of verapamil assessed with quantitated serial treadmill stress testing. American Journal of Cardiology 1981; 48(3):529-535.
62.	Uhlenhuth EH, Glass RM, Fischman MW. Multiple crossover designs with an antianxiety agent and an antidepressant. Psychopharmacology Bulletin 1979; 15(3):37-40.
63.	Parodi O, Maseri A, Simonetti I. Management of unstable angina at rest by verapamil. A double-blind cross-over study in coronary care unit. British Heart Journal 1979; 41(2):167-174.
64.	Gostick NK, Mayhew SR, Million R, Sagar D, Suxena SR, Ingram DF et al. A dose-response study of atenolol in mild to moderate hypertension in general practice. Current Medical Research & Opinion 1977; 5(2):179-184.
65.	Kellner R, Freese ML, Rada RT, Wall FJ. A pilot study of the short-term psychotropic effects of GPA 2640. J Clin Pharmacol 1976; 16(4):194-197.
66.	Huskisson EC, Scott PJ. Floctafenine: a new analgesic for use in rheumatic diseases. Rheumatol Rehabil 1977; 16(1):54-57.
67.	Nikkila M, Inkovaara J, Heikkinen J. Once daily compared with twice daily administration of slow-release diltiazem as monotherapy for hypertension. Ann Med 1991; 23(2):141-145.
68.	Gibelli G, Negrini M, Bruno AM, Fiorini GL, Lambiase M, Magenta G et al. Chronic effects of transdermal nitroglycerin in stable angina pectoris: a within-patient, placebo-controlled study. Int J Clin Pharmacol Ther Toxicol 1989; 27(9):436-441.

Appendix Figure 1: Patients with chronic intestinal pseudo-obstruction treated with prucalopride or placebo for pain relief[14]



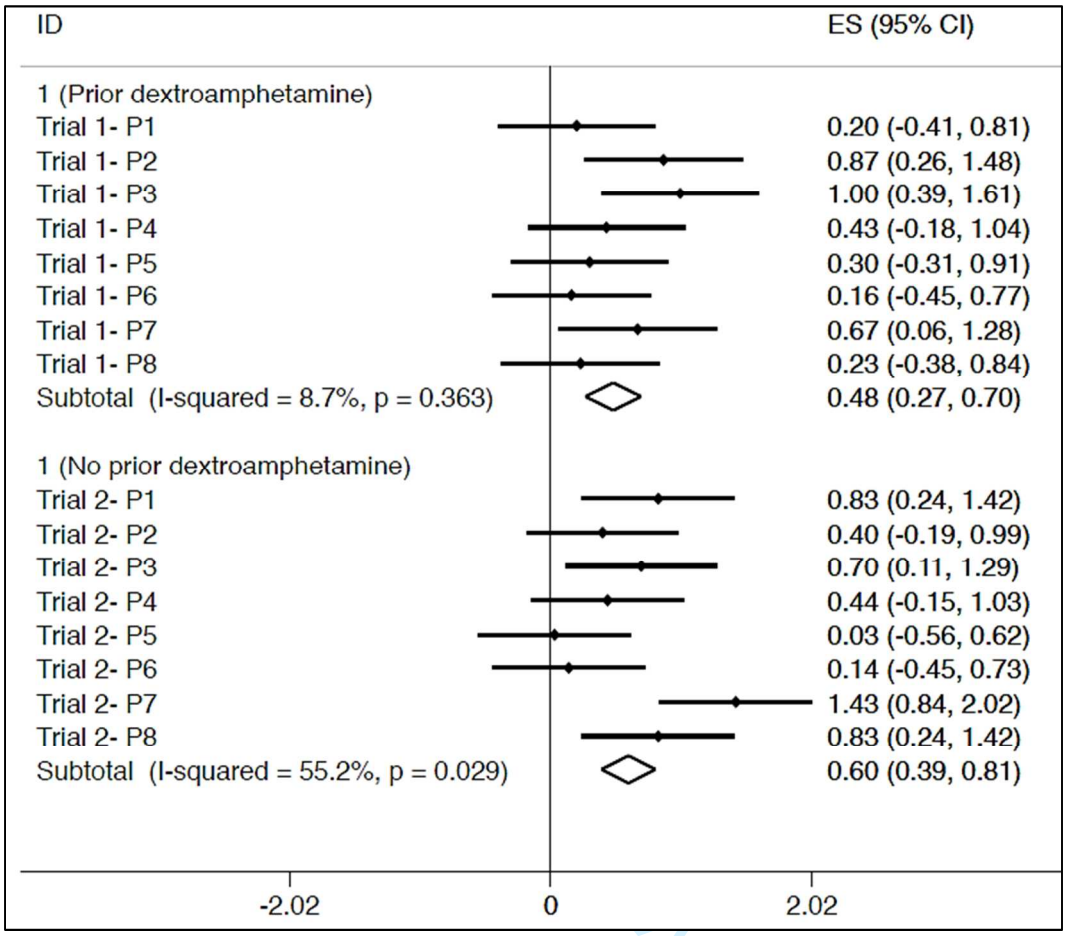
Appendix Figure 1 Legend:
Data from this figure was extracted from the study published by Emmanuel et al in 2011, which investigates the use of prucalopride or placebo for pain relief (among other outcomes) in patients with chronic intestinal pseudo-obstruction. The treatment effect is -0.440 (-0.771 to -0.110) for Appendix Figure 1.

Appendix Figure 2: Patients with chronic tension-type headaches treated with dextroamphetamine or control and effect on mean daily grade decrease in headache[15]



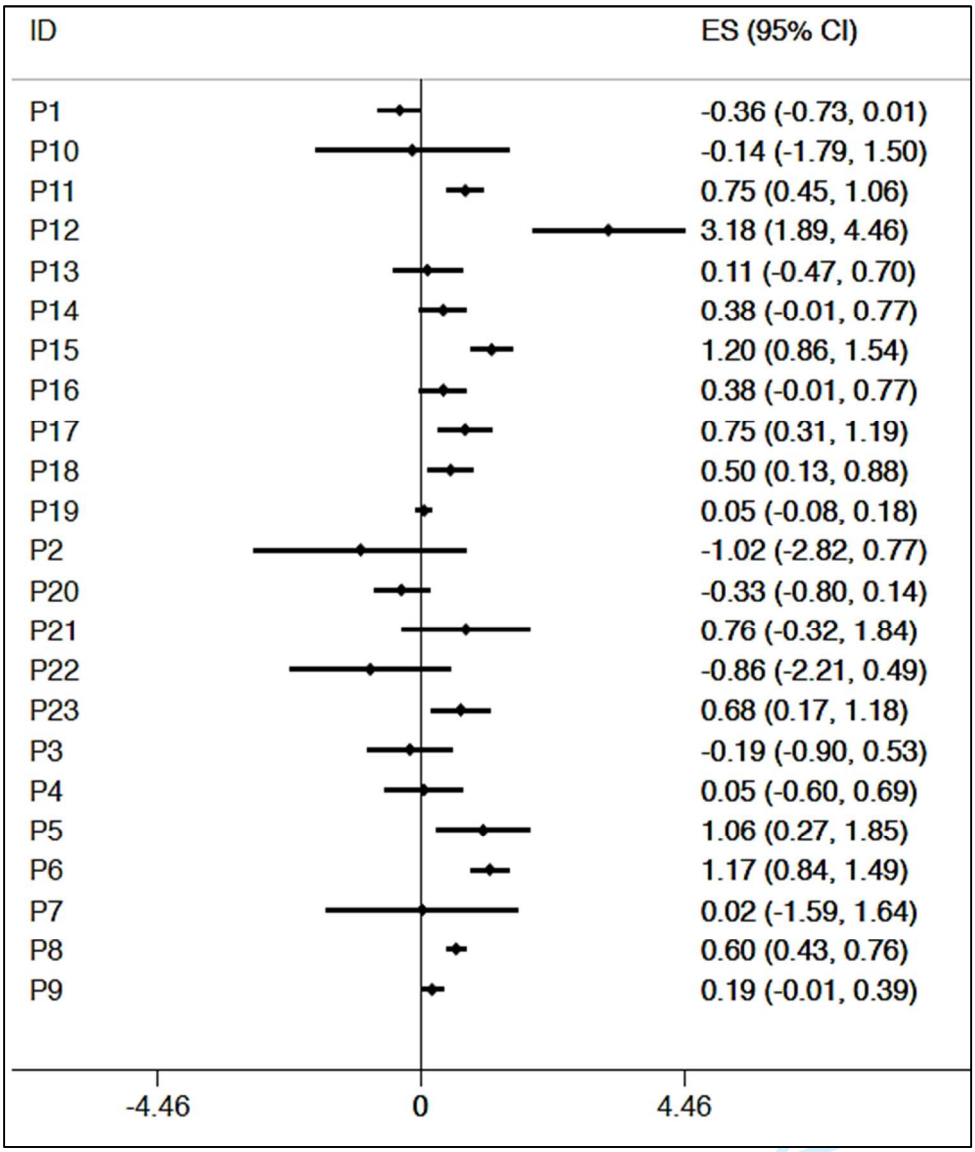
Appendix Figure 2 Legend: Data from this figure was extracted from the study published by Haas et al in 2004, which investigates the use of dextroamphetamine or control in patients with chronic-type for improvement on mean daily grade in headache. The treatment effect is mean daily grade decrease in chronic tension-type headache for Appendix Figure 2.

Appendix Figure 3: Patients with migraine headaches treated with dextroamphetamine or control and effect on mean daily grade decrease in headache[15]



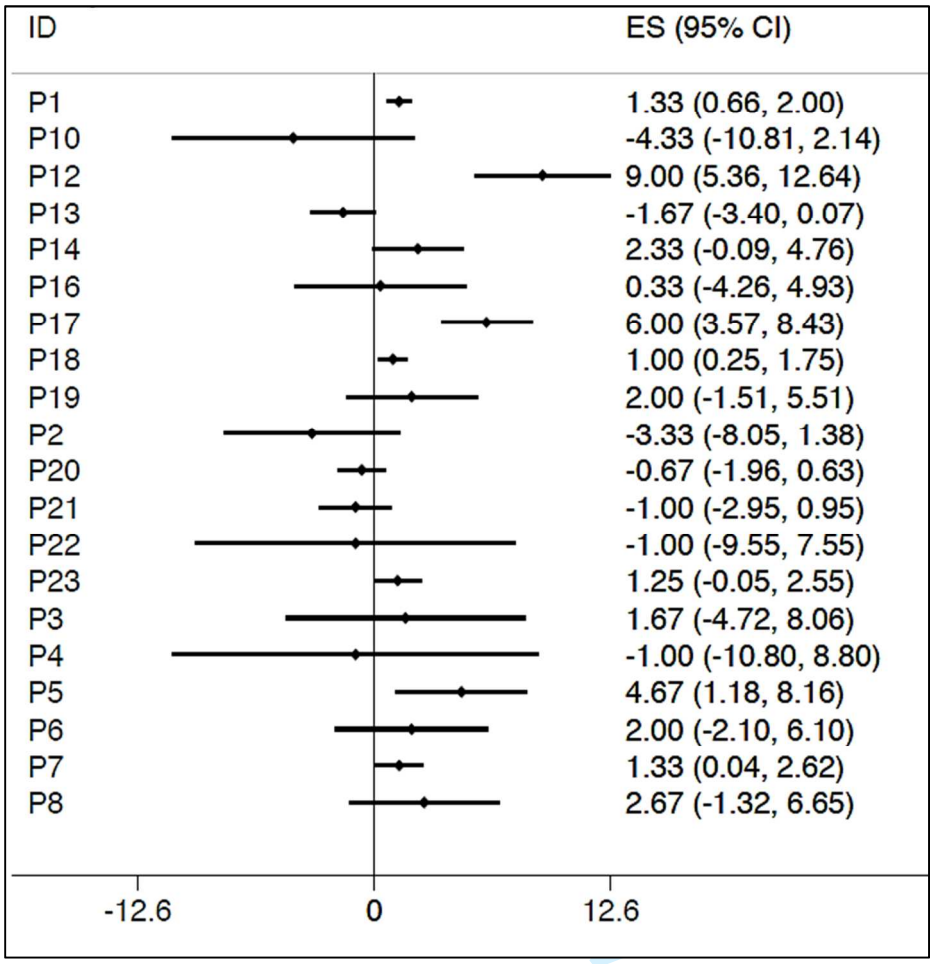
Appendix Figure 3 Legend: Data from this figure was extracted from the study published by Haas et al in 2004, which investigates the use of dextroamphetamine or control in patients with chronic-type and migraine headaches for improvement on mean daily grade in headache. The treatment effect is mean daily grade decrease in migraine headache for Appendix Figure 3.

Appendix Figure 4: Patients with fibromyalgia treated with amitriptyline or placebo and its effect on a 7-point symptom scale[16]



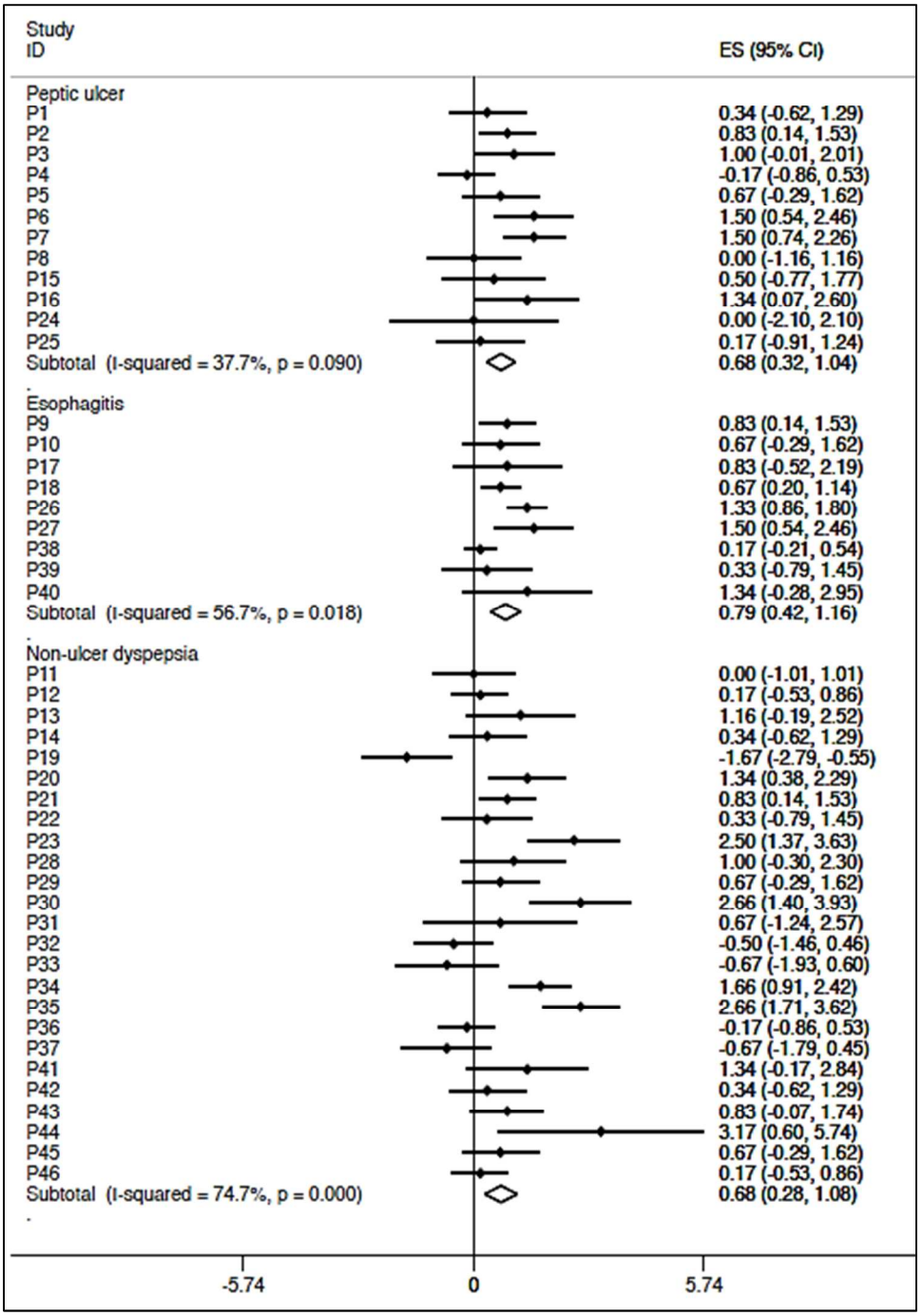
Appendix Figure 4 Legend: Data from this figure was extracted from the study published by Jaeschke et al in 1991, which investigates the effect of amitriptyline or placebo on a 7-point symptom scale in patients with fibromyalgia. The treatment effect is 0.427 (0.210 to 0.645) for Appendix Figure 4.

Appendix Figure 5: Patients with fibromyalgia treated with amitriptyline or placebo and its effect on tender point changes count[16]



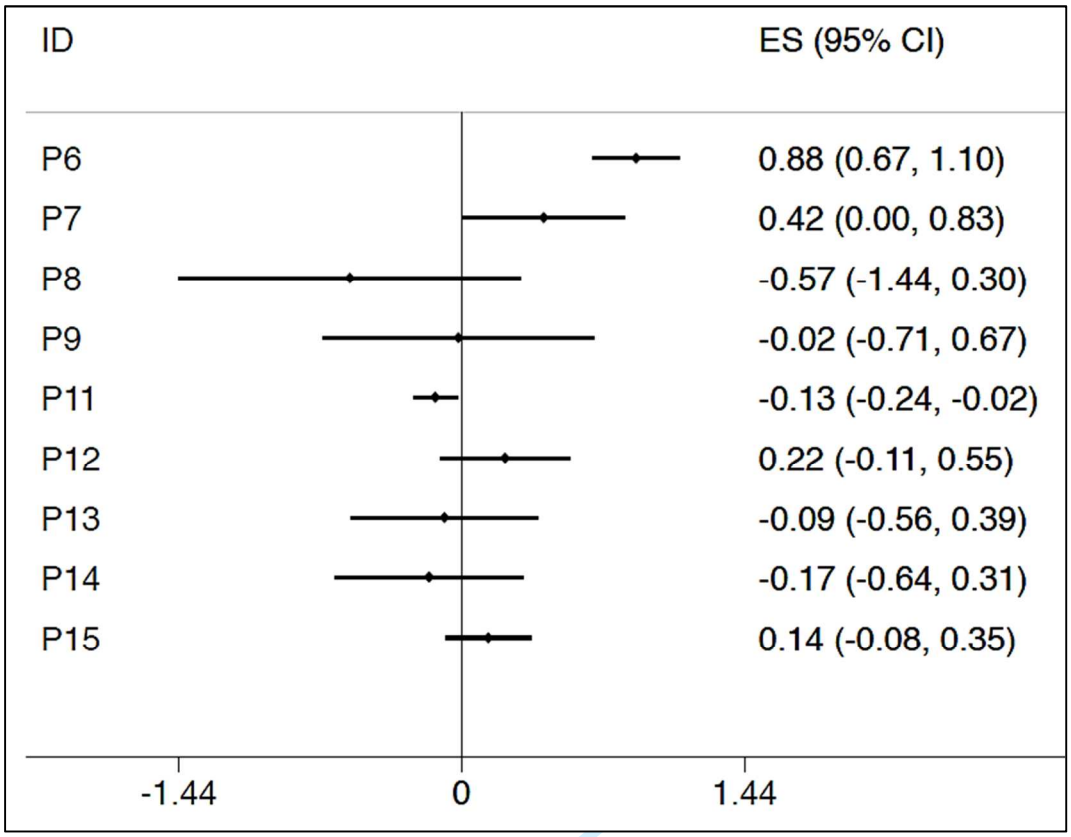
Appendix Figure 5 Legend: Data from this figure was extracted from the study published by Jaeschke et al in 1991, which investigates the effect of amitriptyline or placebo on tender point changes count in patients with fibromyalgia. The treatment effect is 1.320 (0.404 to 2.236) for Appendix Figure 5.

Appendix Figure 6: Patients with peptic ulcers, oesophagitis grade I, II, or III, or with reflux or ulcer-like symptom profiles were treated with cimetidine or placebo and its effect on a 6-point symptom scale[17]



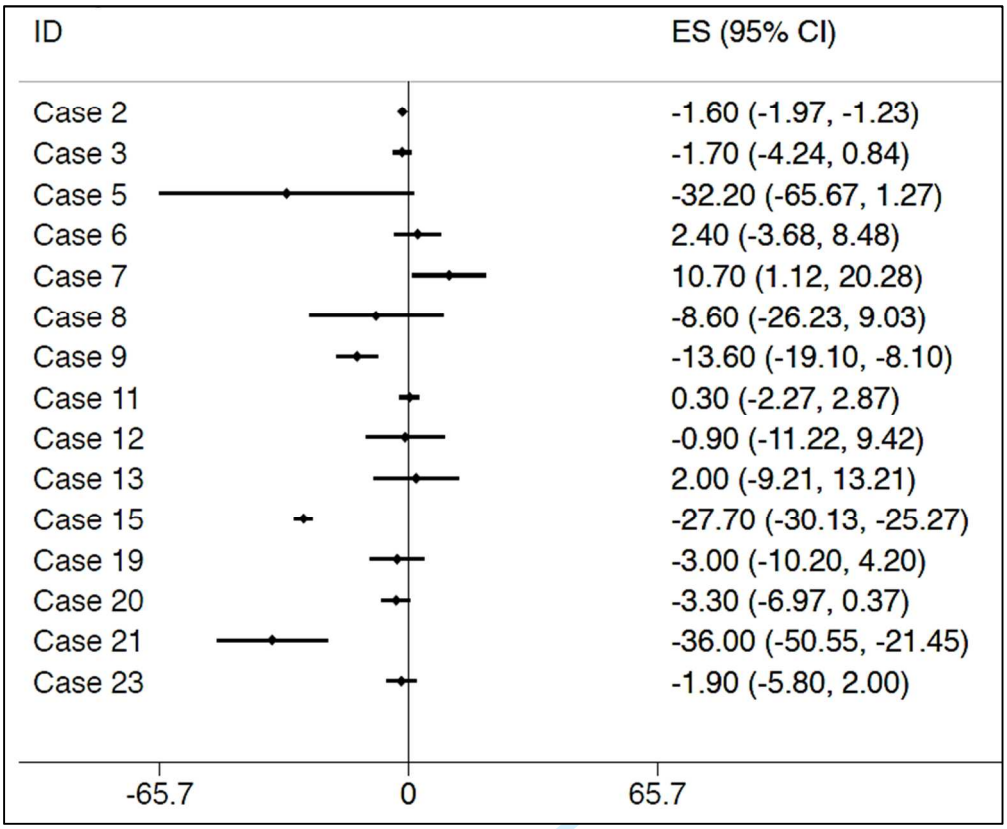
Appendix Figure 6 Legend: Data from this figure was extracted from the study published by Johannessen et al in 1992, which investigates the effect of cimetidine or placebo on a 6-point symptom scale in patients with peptic ulcers, oesophagitis grade I, II, or III, or with reflux or ulcer-like symptom profiles. The treatment effect is 0.698 (0.466 to 0.931) for Appendix Figure 6.

Appendix Figure 7: Patients with irreversible chronic airflow limitation treated with theophylline or placebo and its effect on dyspnea[18]



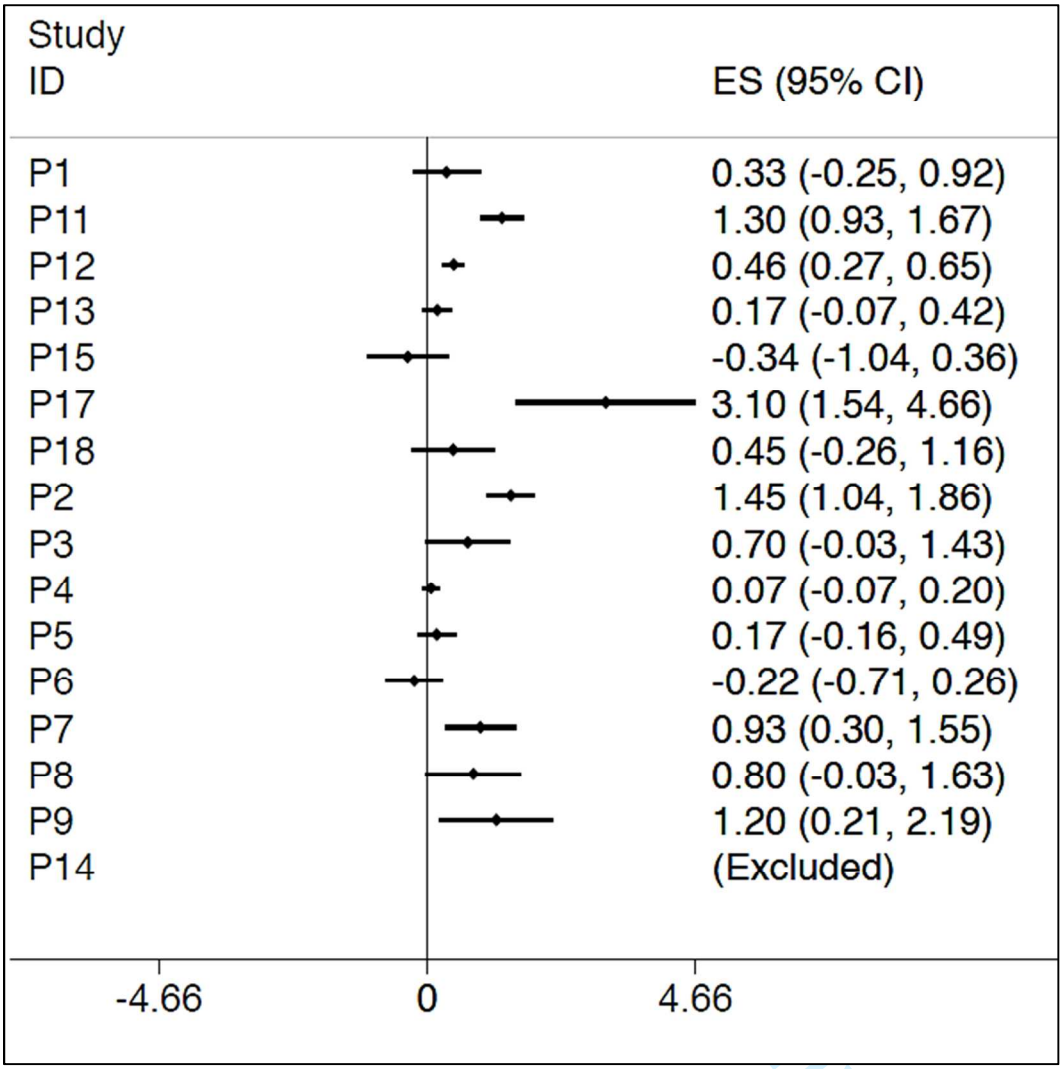
Appendix Figure 7 Legend: Data from this figure was extracted from the study published by Mahon et al in 1996, which investigates the effect of theophylline or placebo on dyspnea in patients with irreversible chronic airflow limitation. The treatment effect is 0.125 (-0.181 to 0.430) for Appendix Figure 7.

Appendix Figure 8: Patients with osteoarthritic pain treated with paracetmol and diclofenac and its effect on stiffness[19]



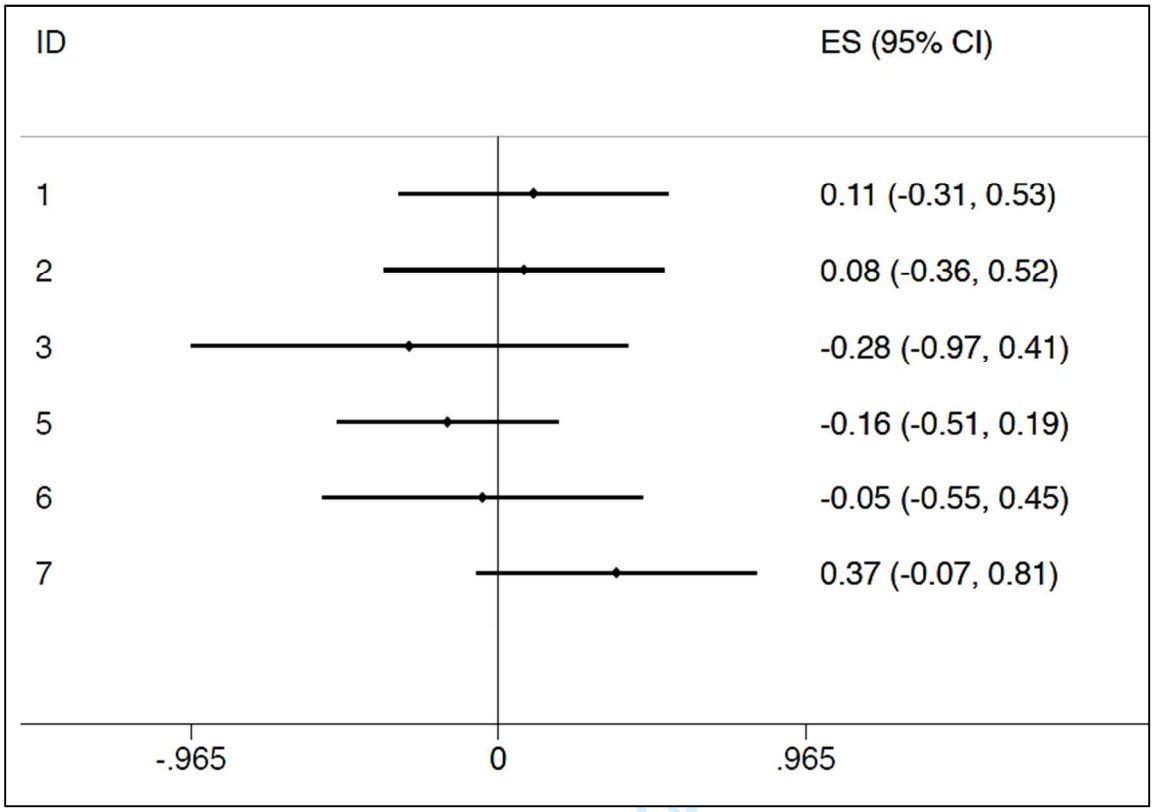
Appendix Figure 8 Legend: Data from this figure was extracted from the study published by March et al in 1994, which investigates the effect of paracetmol and diclofenac on stiffness in patients with osteoarthritic pain. The treatment effect is mean difference in stiffness (mm) for Appendix Figure 8.

Appendix Figure 9: Patients with nonreversible chronic airflow limitation treated with either ipratropium bromide, theophylline, salbutamol, or beclomethane (all compared to placebo) and its effect on a 4-item symptom questionnaire[20]



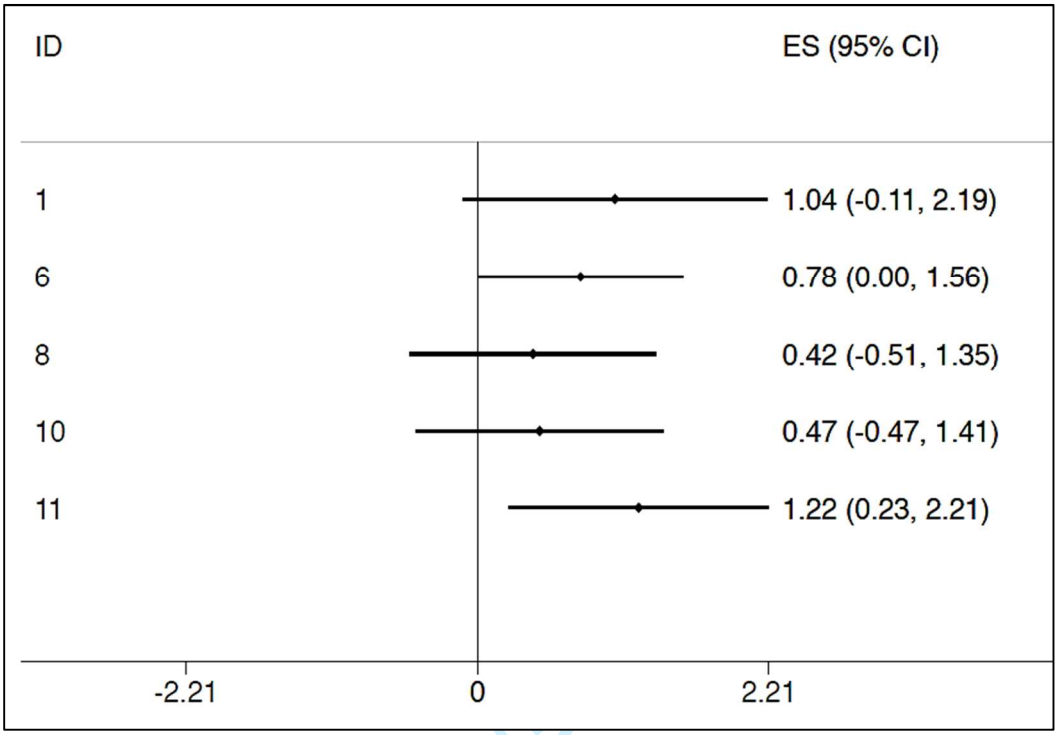
Appendix Figure 9 Legend: Data from this figure was extracted from the study published by Patel et al in 1991, which investigates the effect of ipratropium bromide, theophylline, salbutamol, or beclomethane (all compared to placebo) on a 4-item symptom questionnaire in patients with nonreversible chronic airflow limitation. The treatment effect is 0.240 (0.131 to 0.350) for Appendix Figure 9.

Appendix Figure 10: Patients previously taking warfarin for either atrial fibrillation or deep vein thrombosis treated with apo-warfarin and 20coumadin and its effect on international normalized ratio[12]



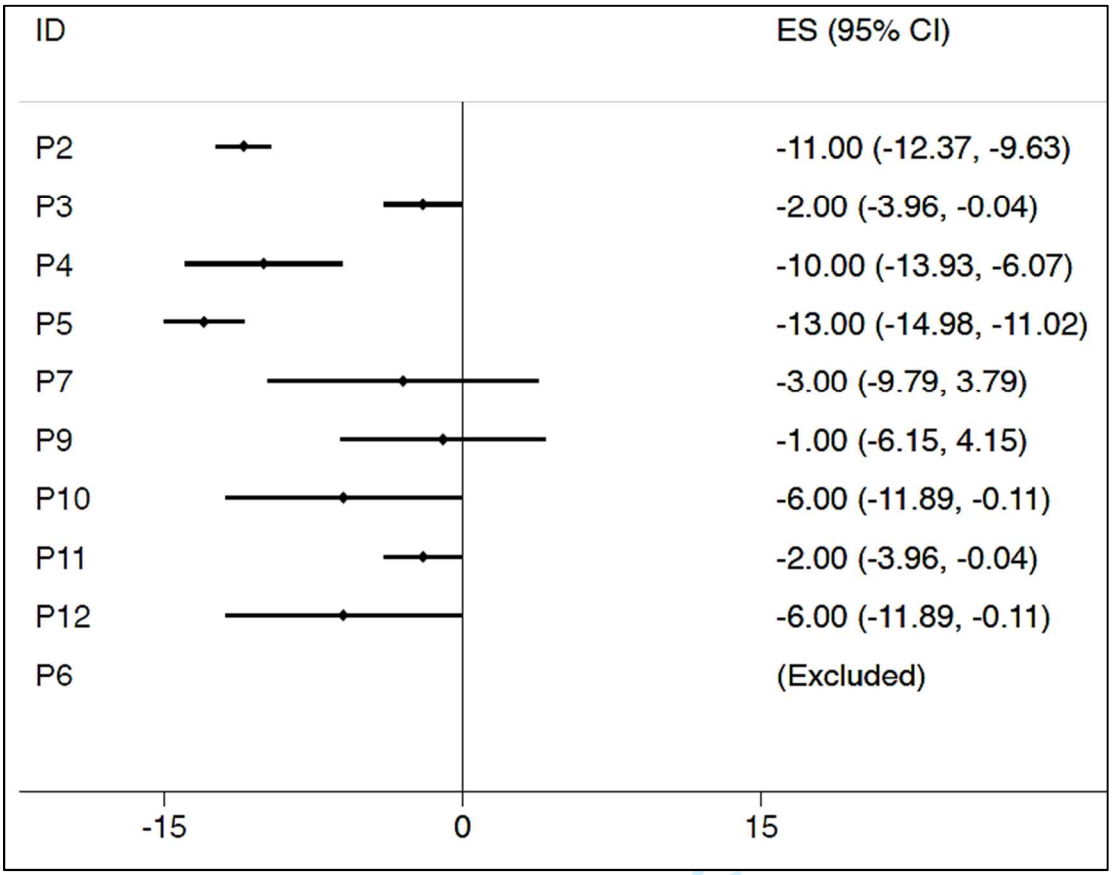
Appendix Figure 10 Legend: Data from this figure was extracted from the study published by Pereira et al in 1995, which investigates the effect of apo-warfarin and Coumadin on international normalized ratio in patients previously taking warfarin for either atrial fibrillation or deep vein thrombosis. The treatment effect is 0.027 (-0.155 to 0.209) for Appendix Figure 10.

Appendix Figure 11: Hospitalized children and adolescents with attention-deficit hyperactivity disorder treated with methylphenidate and placebo and its effect on Conners 15-item rating scale scores[21]



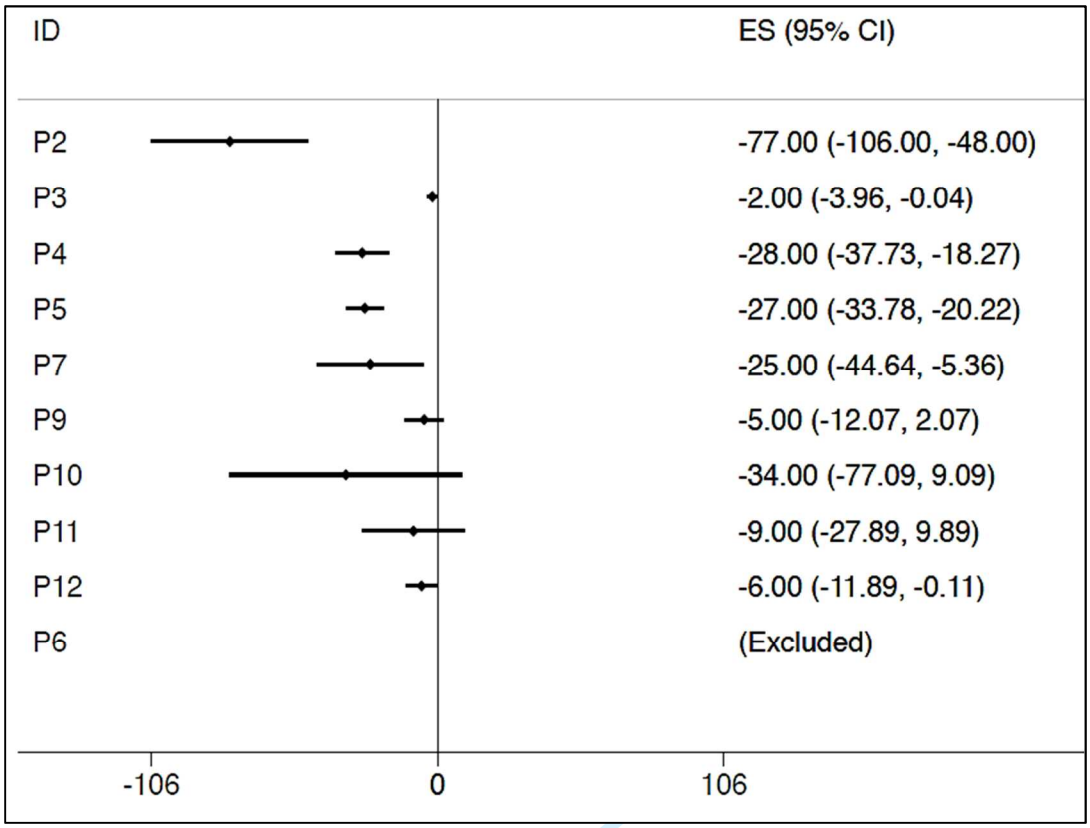
Appendix Figure 11 Legend: Data from this figure was extracted from the study published by Wallace et al in 1994, which investigates the effect of methylphenidate and placebo on Conners 15-item rating scale scores in hospitalized children and adolescents with attention-deficit hyperactivity disorder. The treatment effect is 0.759 (0.341 to 1.178) for Appendix Figure 11.

Appendix Figure 12: Patients already prescribed quinine treated with quinine sulphate and placebo, and its effect on changes in number of cramps[22]



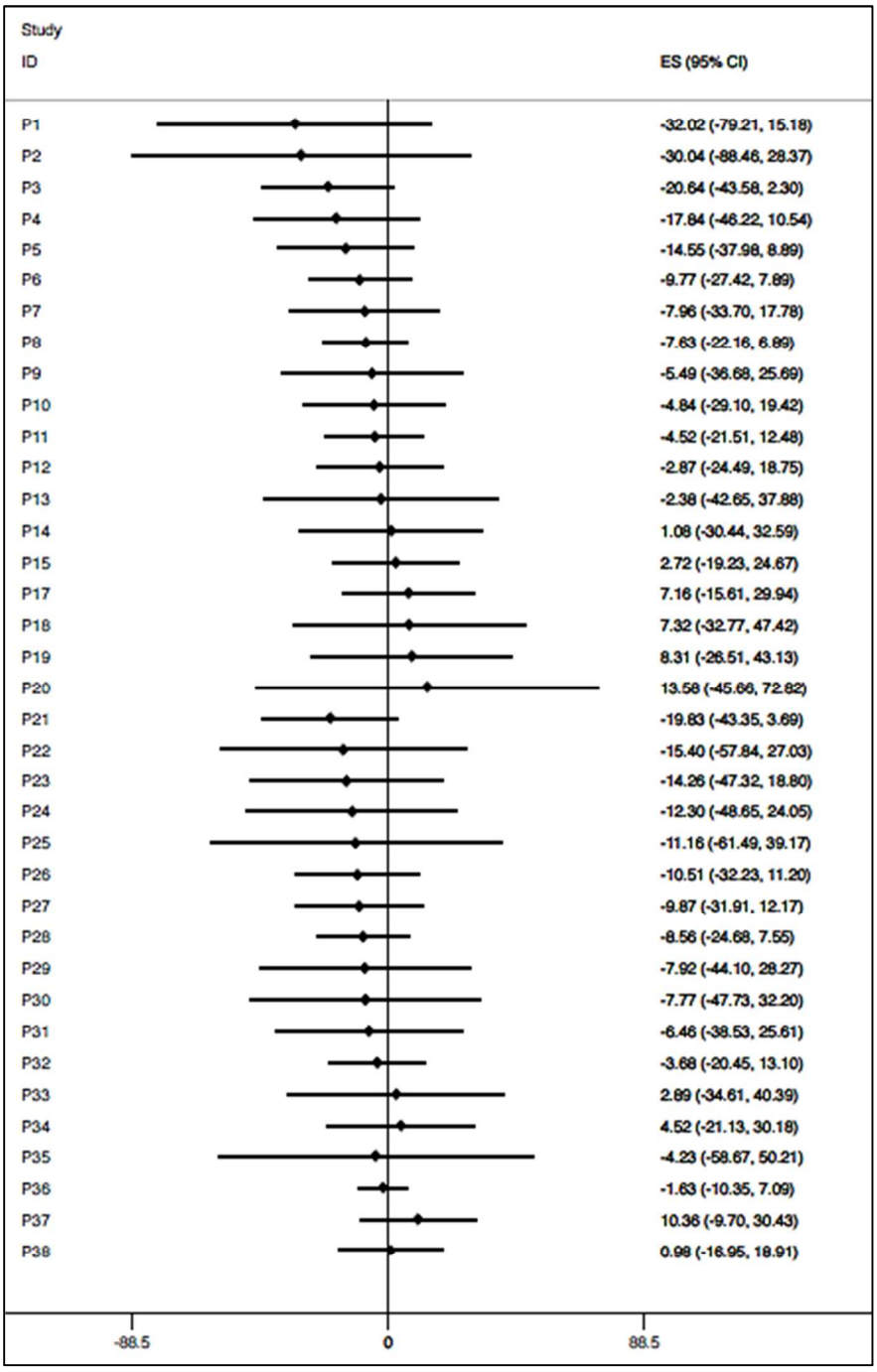
Appendix Figure 12 Legend: Data from this figure was extracted from the study published by Woodfield et al in 2005, which investigates the effect of quinine sulphate and placebo on changes in number of cramps in patients already prescribed quinine. The treatment effect is -18.823 (-28.527 to -9.120) for Appendix Figure 12.

Appendix Figure 13: Patients already prescribed quinine treated with quinine sulphate and placebo, and its effect on total days with cramps[22]



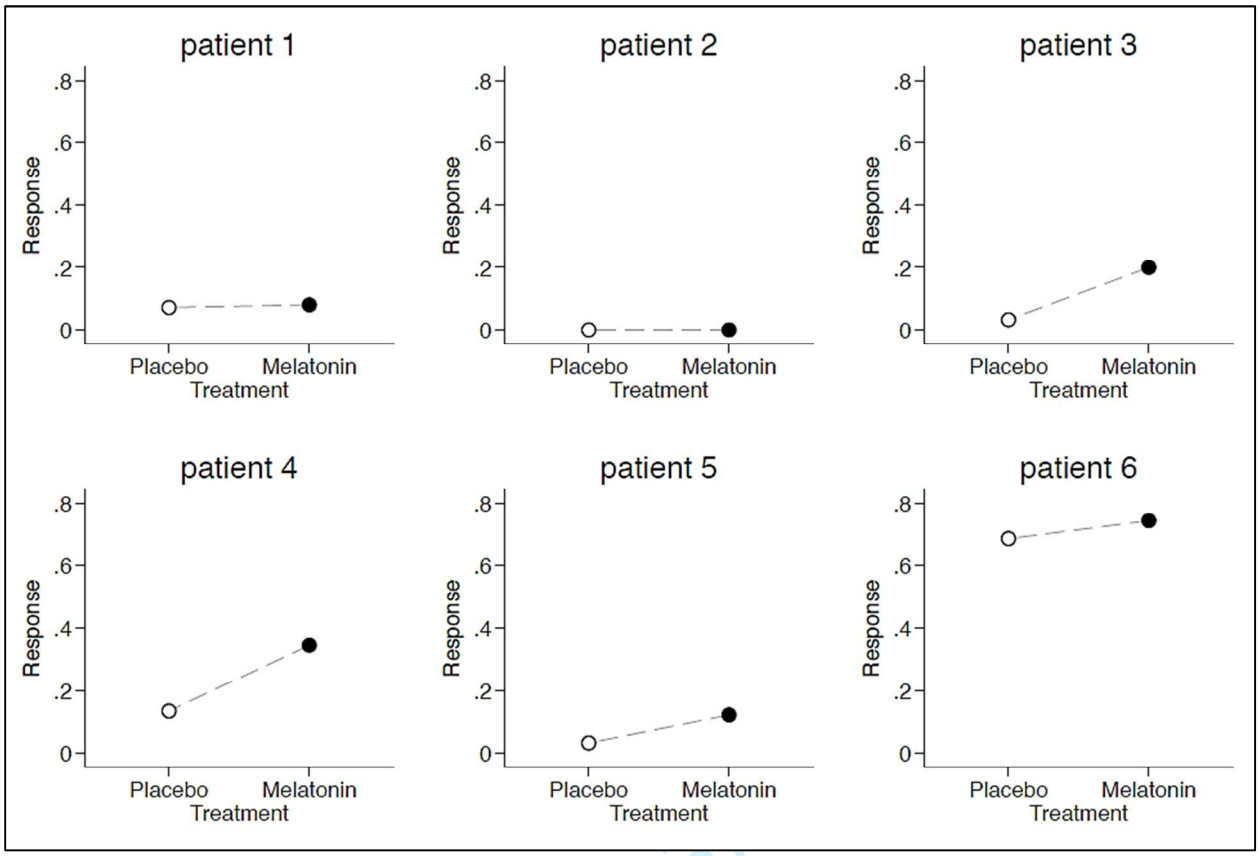
Appendix Figure 13 Legend: Data from this figure was extracted from the study published by Woodfield et al in 2005, which investigates the effect of quinine sulphate and placebo on total days with cramps in patients already prescribed quinine. The treatment effect is -6.181 (-9.798 to -2.563) for Appendix Figure 13.

Appendix Figure 14: Patients with fibromyalgia syndrome treated with amitriptyline and the combination amitriptyline and fluoxetine and its effect on the Fibromyalgia Impact Questionnaire[23]



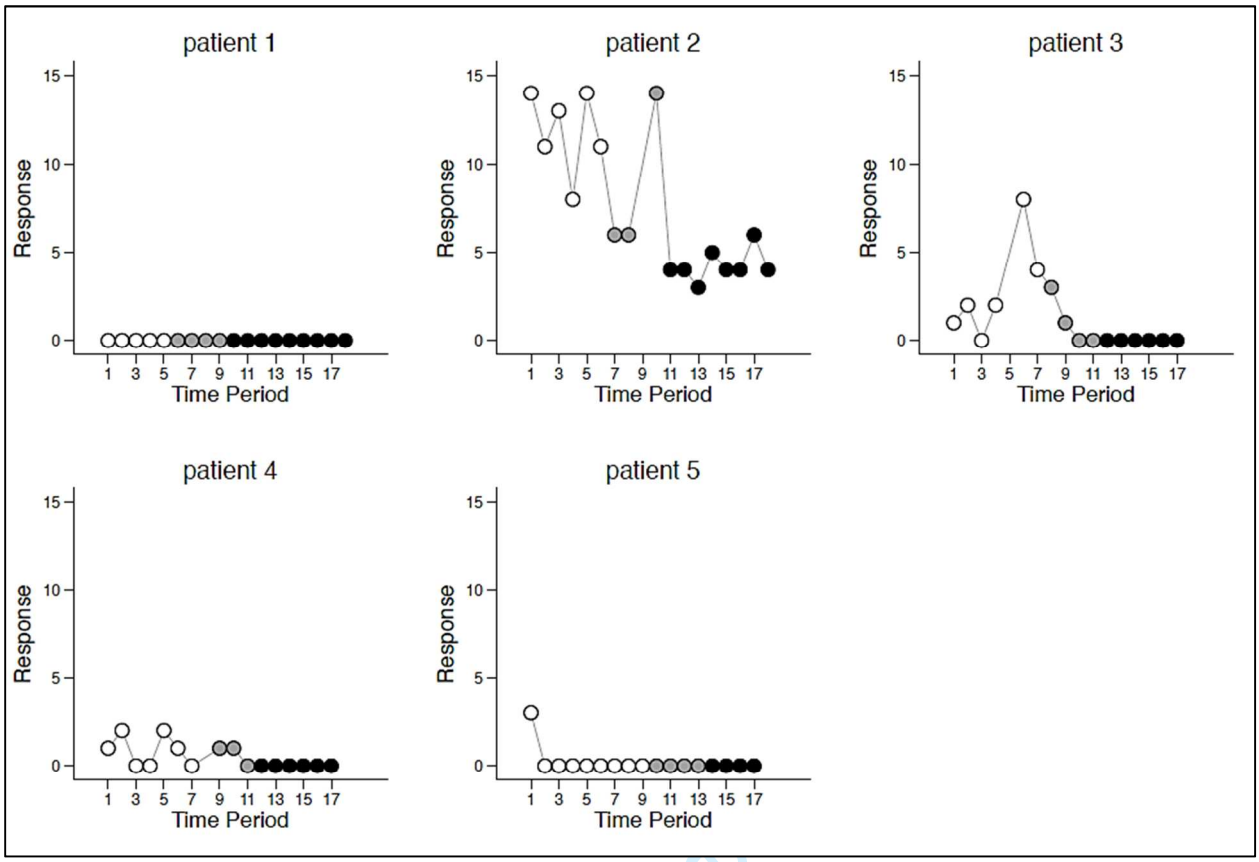
Appendix Figure 14 Legend: Data from this figure was extracted from the study published by Zucker et al in 2006, which investigates the effect of amitriptyline and the combination amitriptyline and fluoxetine on Fibromyalgia Impact Questionnaire in patients with fibromyalgia syndrome. The treatment effect is -5.019 (-8.784 to -1.254) for Appendix Figure 14.

Appendix Figure 15: Children with mental retardation and fragmented sleep treated with melatonin and placebo and its effect on nights without awakening[1]



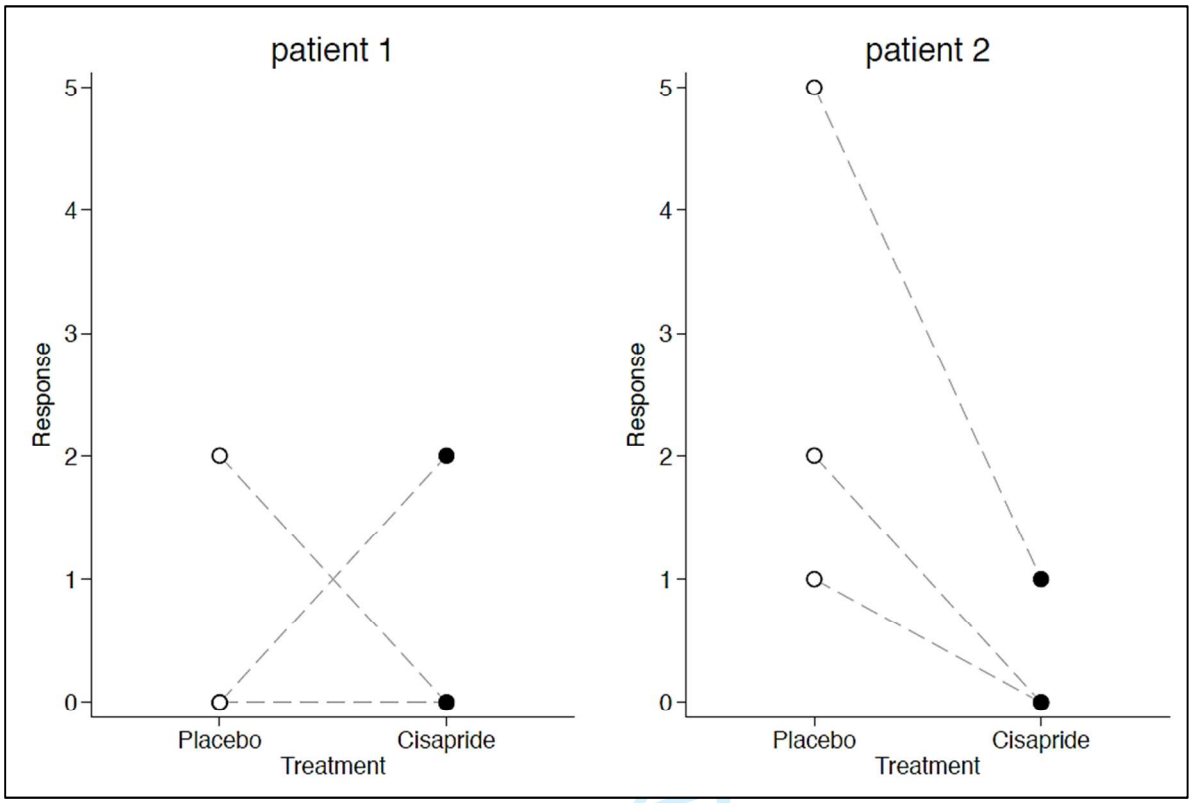
Appendix Figure 15 Legend: Data from this figure was extracted from the study published by Camfield et al in 1996, which investigates the effect of melatonin and placebo on nights without awakening in children with mental retardation and fragmented sleep. The treatment effect is 0.84 (0.20 to 1.48) for Appendix Figure 15. White circles indicate placebo; black circles indicate melatonin.

Appendix Figure 16: Patients with traumatic spinal cord lesions treated with baclofen and placebo and its effect on anxiety[2]



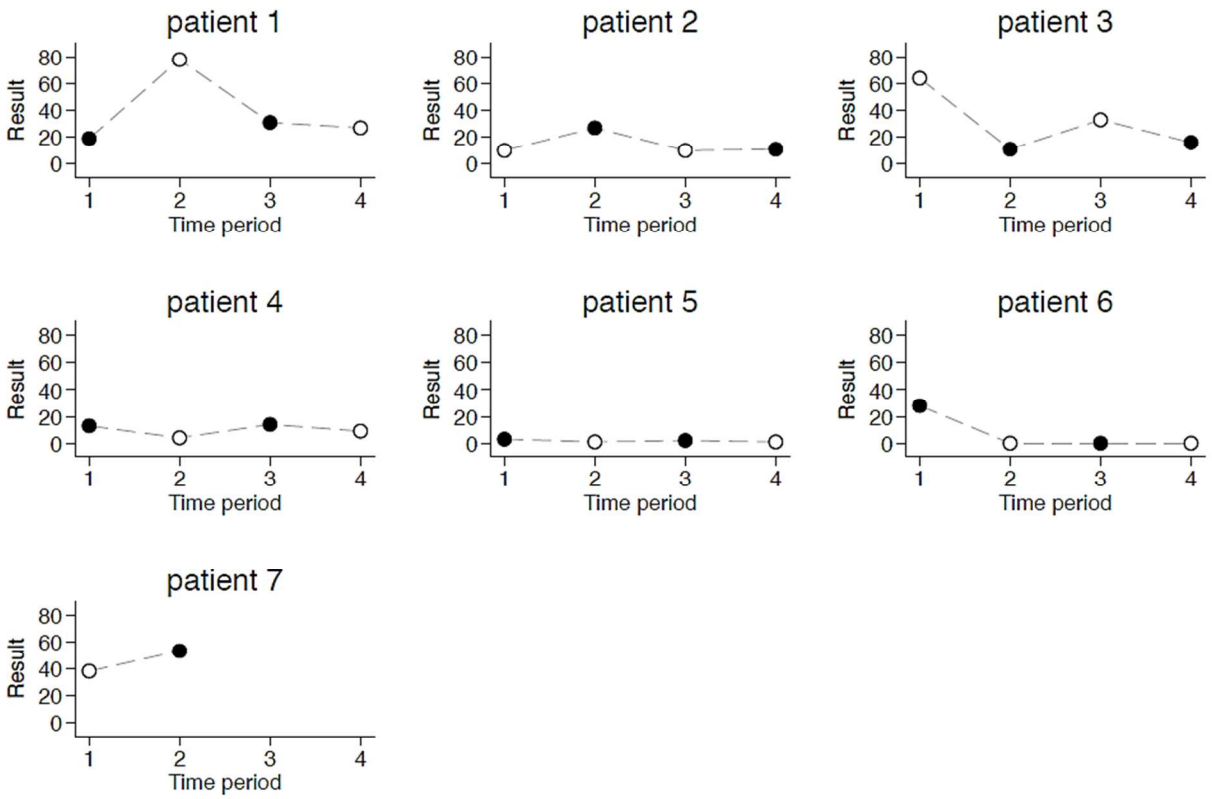
Appendix Figure 16 Legend: Data from this figure was extracted from the study published by Hinderer et al in 1990, which investigates the effect of baclofen and placebo on anxiety in patients with traumatic spinal cord lesions. The treatment effect is -1.06 (-1.88 to -0.23) for Appendix Figure 16. White circles indicate placebo; grey circles indicate a half dose (40 mg/day) of baclofen; black circles indicate a full dose (80 mg/day) of baclofen.

Appendix Figure 17: Children with gastroesophageal reflux treated with cisapride and placebo and its effect on emetic episodes per day[3]



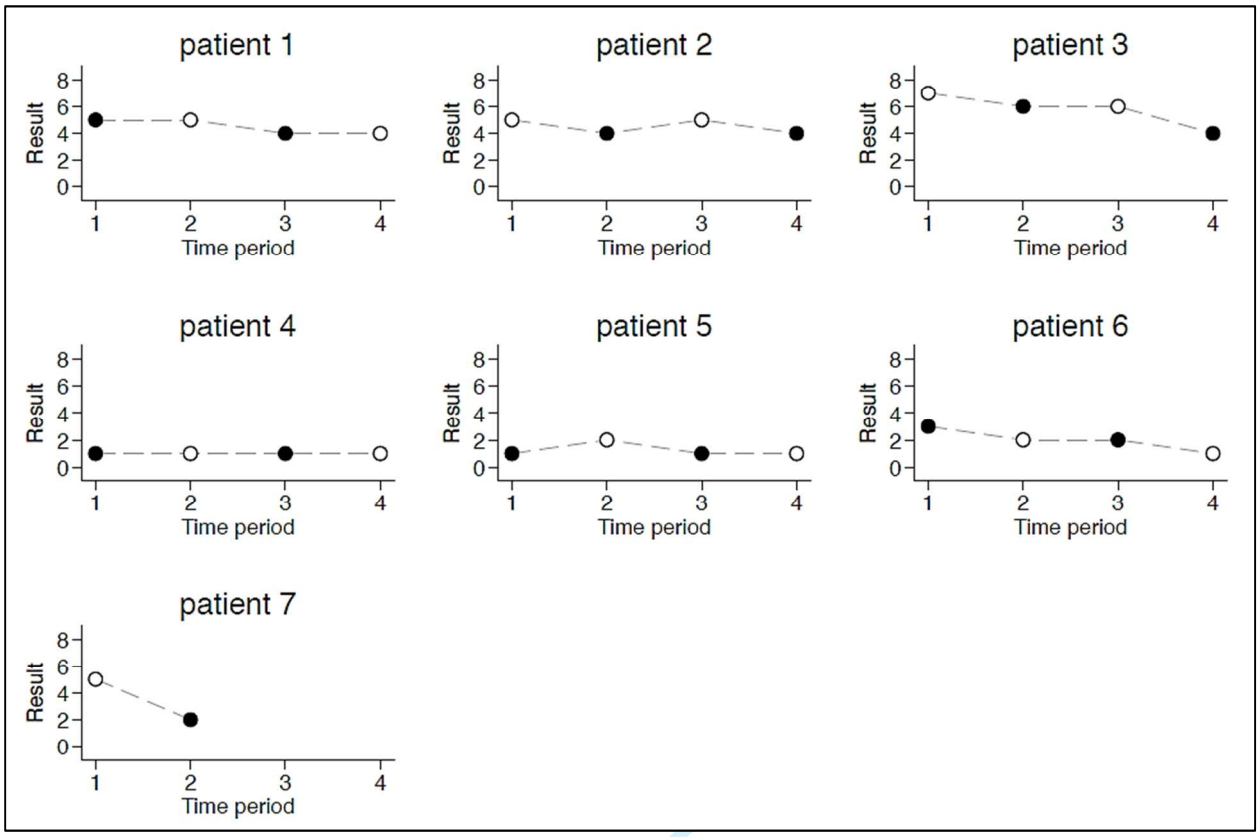
Appendix Figure 17 Legend: Data from this figure was extracted from the study published by Langer et al in 1993, which investigates the effect of cisapride and placebo on emetic episodes per day in children with gastroesophageal reflux. The treatment effect is -1.20 (-2.49 to 0.09) for Appendix Figure 17. White circles indicate placebo; black circles indicate cisapride.

Appendix Figure 18: Nonsmokers with ulcerative colitis treated with nicotine gum and placebo and its effect on abdominal pain[4]



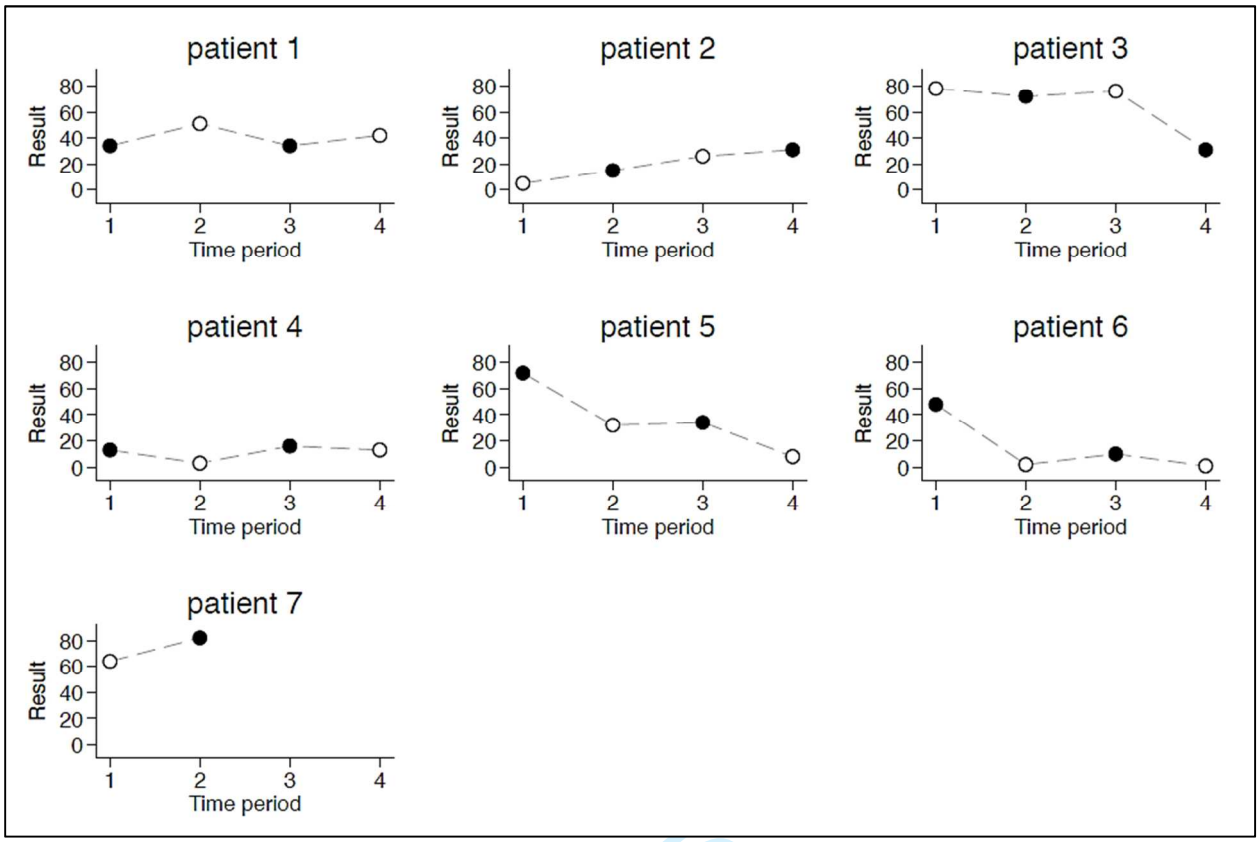
Appendix Figure 18 Legend: Data from this figure was extracted from the study published by Lashner et al in 1990, which investigates the effect of nicotine gum and placebo on abdominal pain in nonsmokers with ulcerative colitis. The treatment effect is -3.62 (-15.84 to 8.61) for Appendix Figure 18. White circles indicate placebo gum; black circles indicate nicotine gum.

Appendix Figure 19: Nonsmokers with ulcerative colitis treated with nicotine gum and placebo and its effect on bowel movements per day[4]



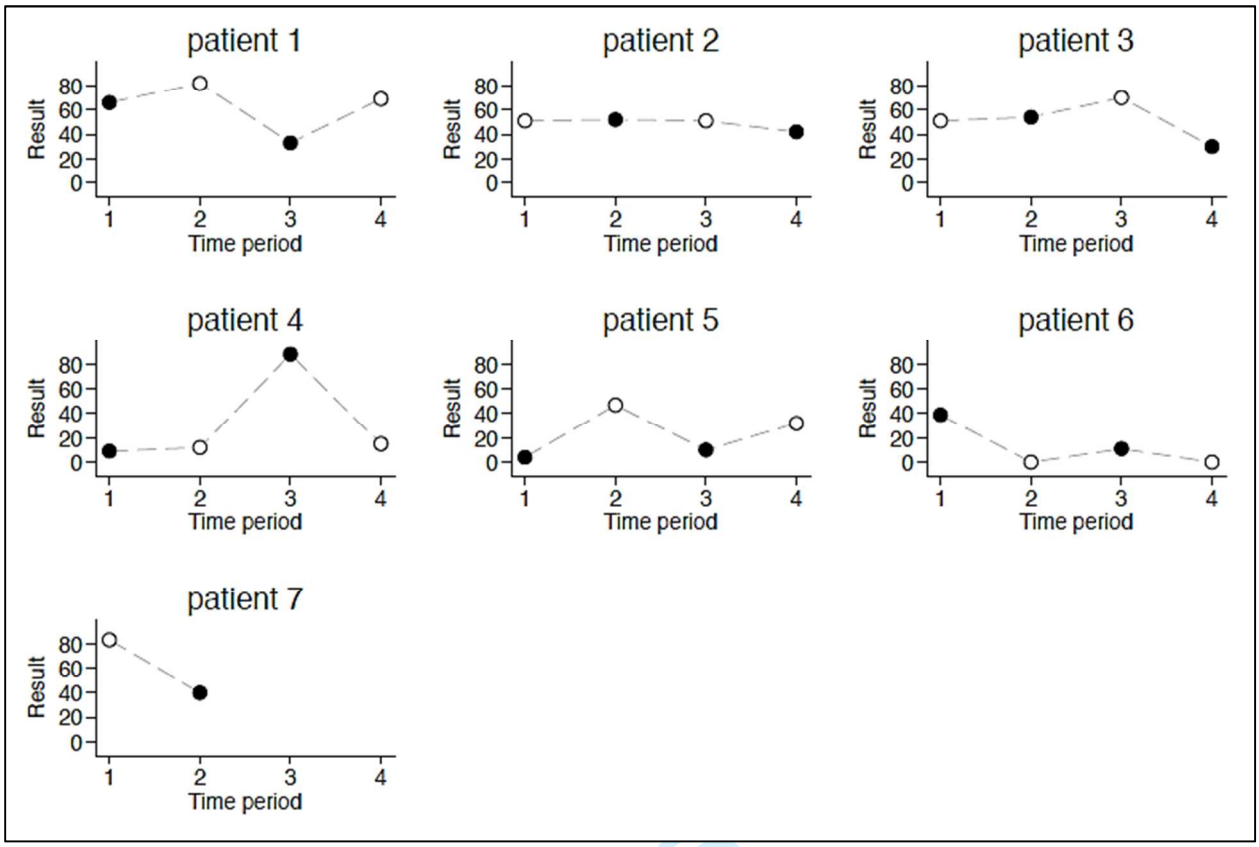
Appendix Figure 19 Legend: Data from this figure was extracted from the study published by Lashner et al in 1990, which investigates the effect of nicotine gum and placebo on bowel movements per day in nonsmokers with ulcerative colitis. The treatment effect is -0.56 (-1.22 to 0.09) for Appendix Figure 19. White circles indicate placebo gum; black circles indicate nicotine gum.

Appendix Figure 20: Nonsmokers with ulcerative colitis treated with nicotine gum and placebo and its effect on consistency of bowel movements[4]



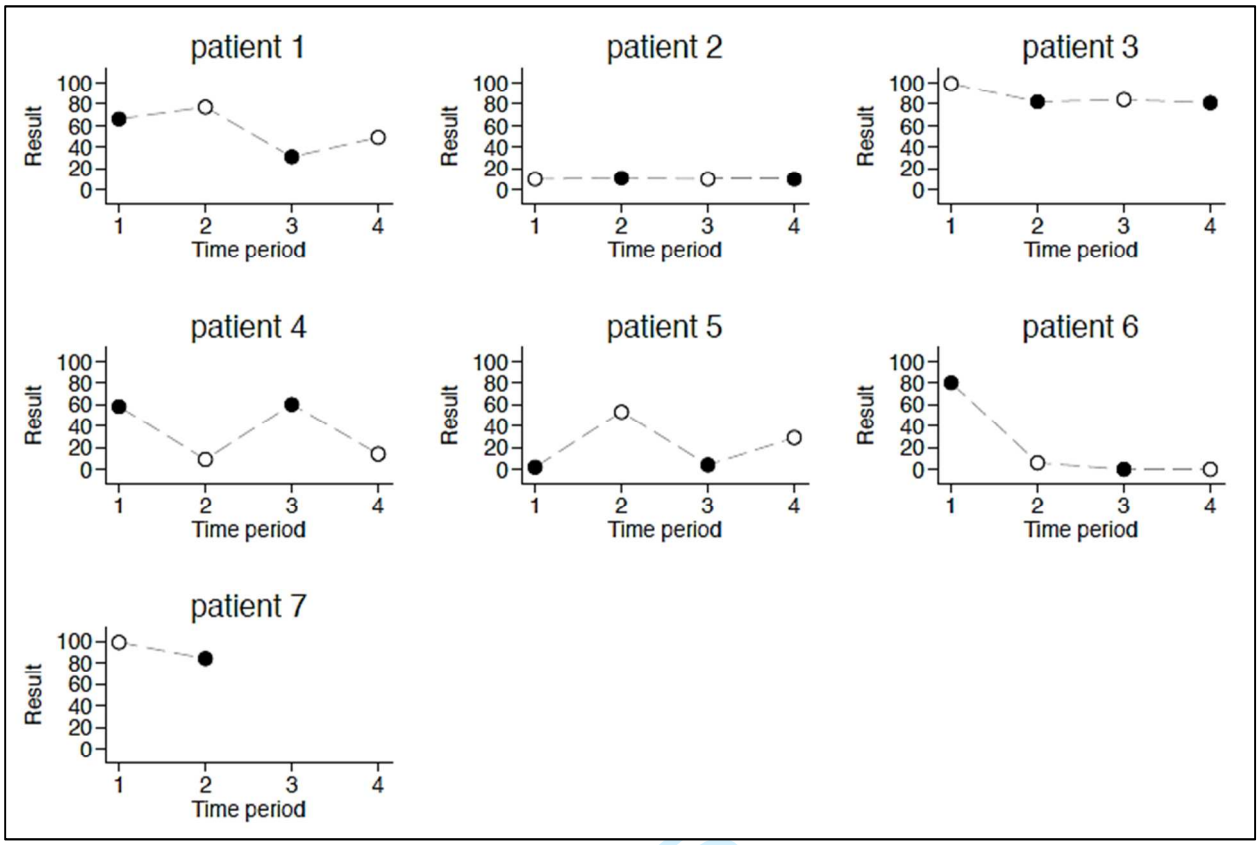
Appendix Figure 20 Legend: Data from this figure was extracted from the study published by Lashner et al in 1990, which investigates the effect of nicotine gum and placebo on consistency of bowel movements in nonsmokers with ulcerative colitis. The treatment effect is 7.00 (-6.29 to 20.29) for Appendix Figure 20. White circles indicate placebo gum; black circles indicate nicotine gum.

Appendix Figure 21: Nonsmokers with ulcerative colitis treated with nicotine gum and placebo and its effect on general sense of well-being[4]



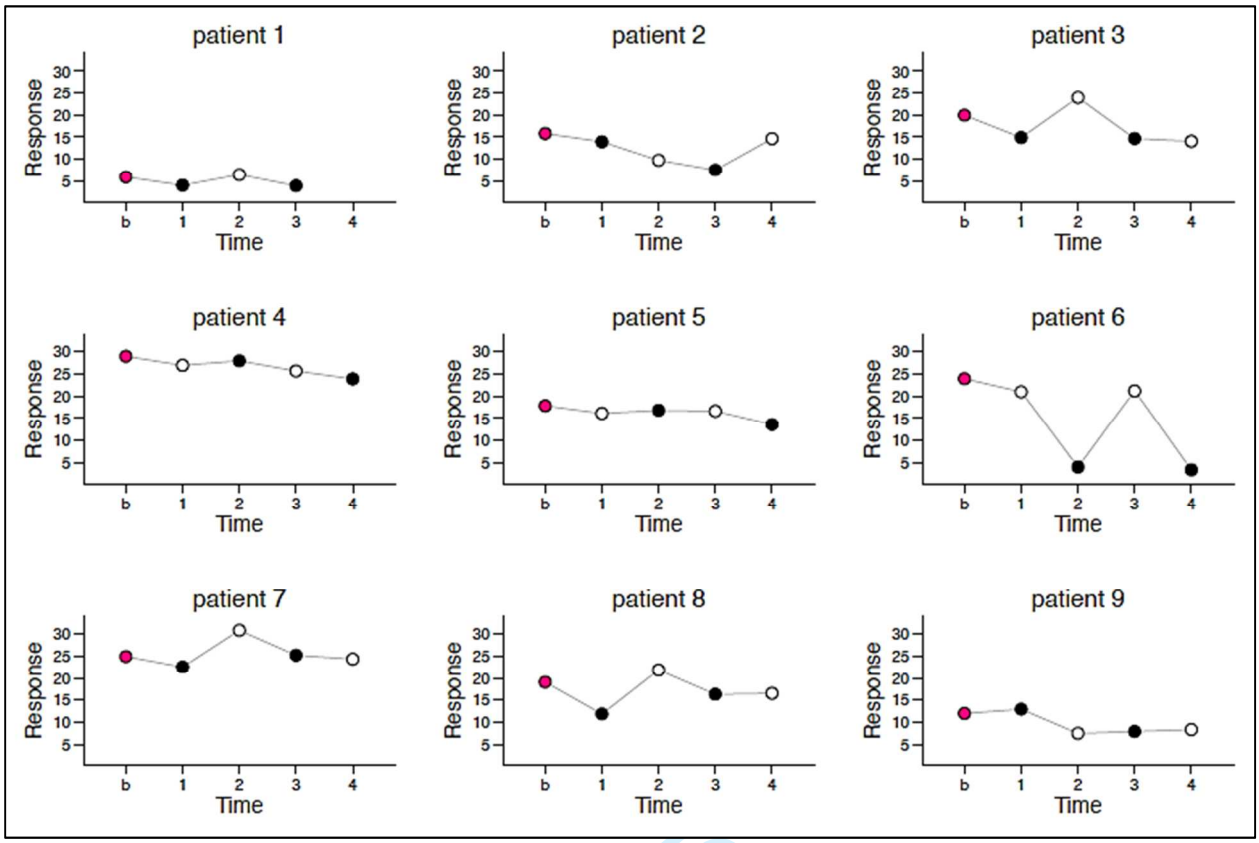
Appendix Figure 21 Legend: Data from this figure was extracted from the study published by Lashner et al in 1990, which investigates the effect of nicotine gum and placebo on general sense of well-being in nonsmokers with ulcerative colitis. The treatment effect is -6.54 (-23.62 to 10.56) for Appendix Figure 21. White circles indicate placebo gum; black circles indicate nicotine gum.

Appendix Figure 22: Nonsmokers with ulcerative colitis treated with nicotine gum and placebo and its effect on hematochezia[4]



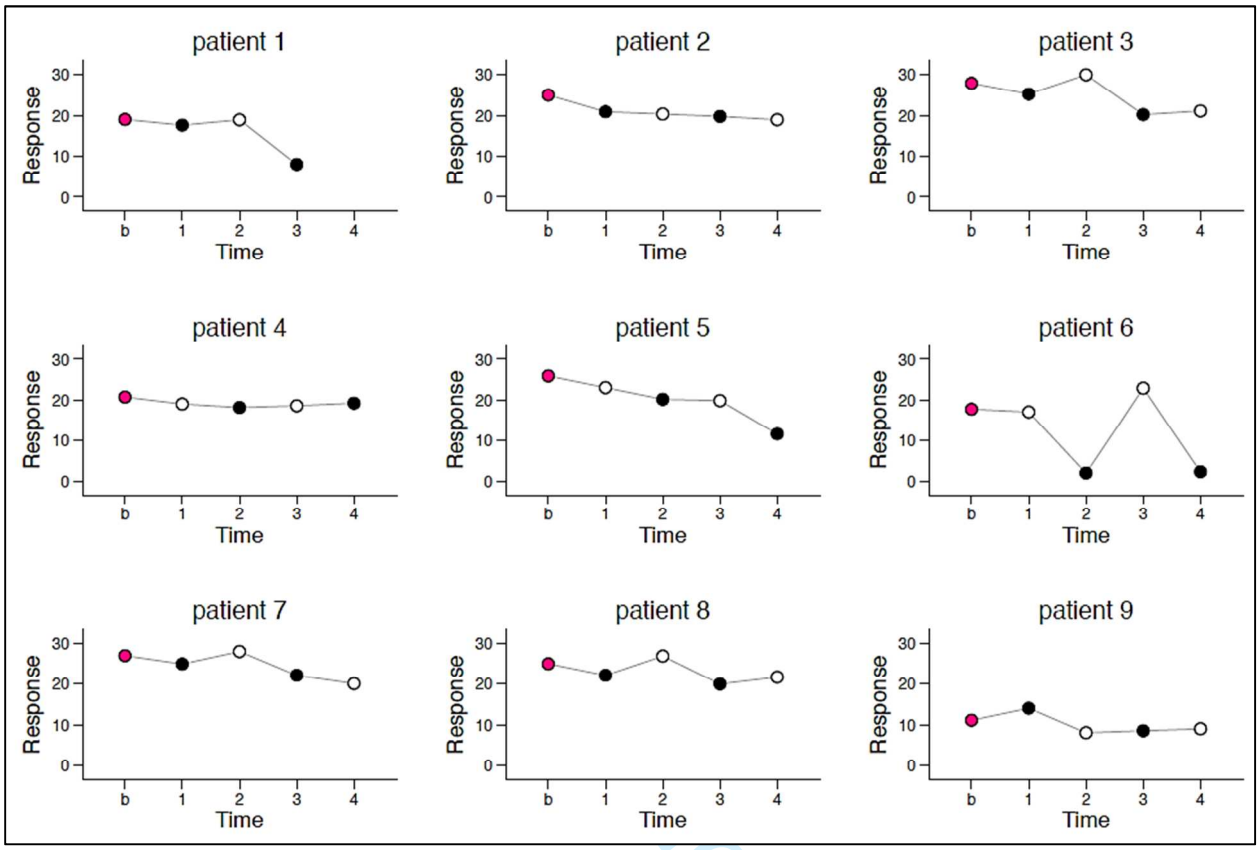
Appendix Figure 22 Legend: Data from this figure was extracted from the study published by Lashner et al in 1990, which investigates the effect of nicotine gum and placebo on hematochezia in nonsmokers with ulcerative colitis. The treatment effect is 2.35 (-17.21 to 21.90) for Appendix Figure 22. White circles indicate placebo gum; black circles indicate nicotine gum.

Appendix Figure 23: Patients with chronic depression and a diagnosis of major depression or dysthymia treated with sulpiride and placebo and its effect on anxiety[5]



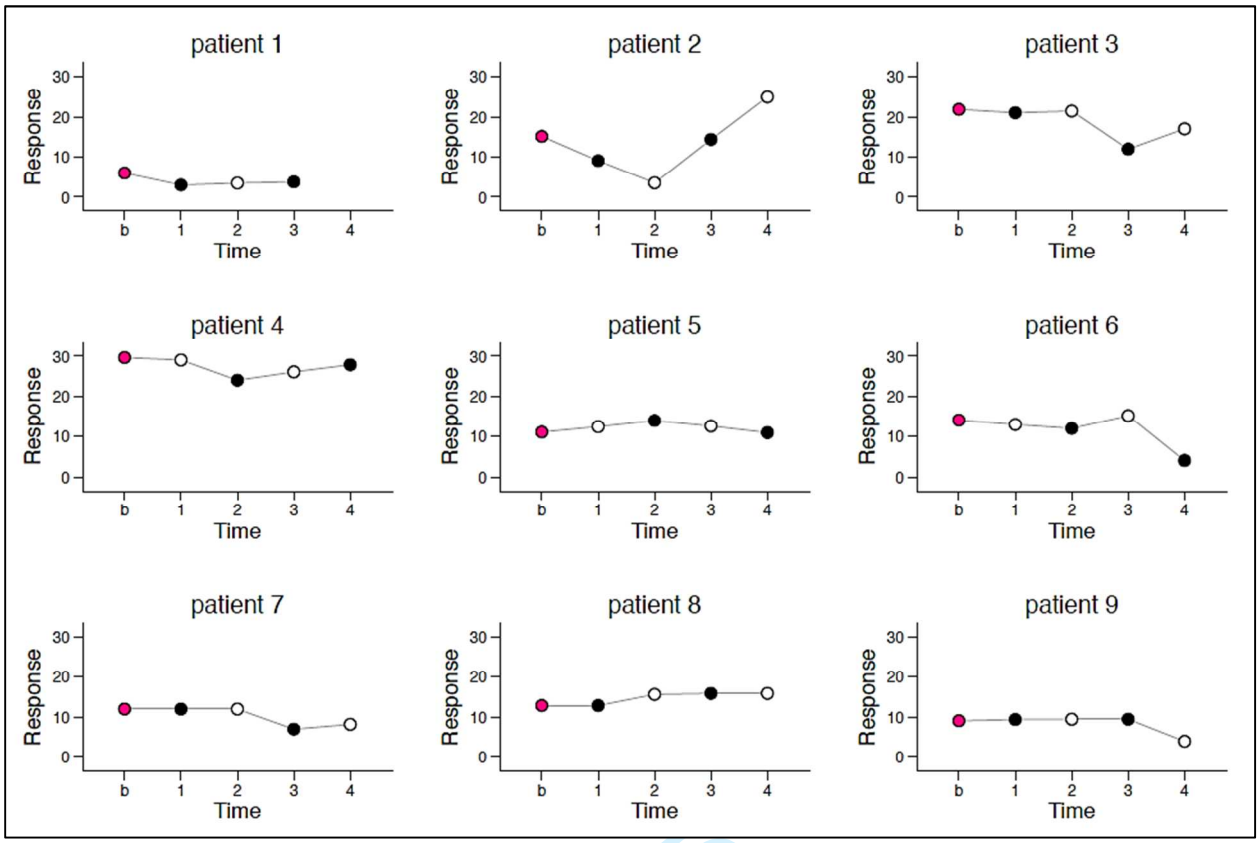
Appendix Figure 23 Legend: Data from this figure was extracted from the study published by Maier et al in 1994, which investigates the effect of sulpiride and placebo on anxiety in patients with chronic depression and a diagnosis of major depression or dysthymia. The treatment effect is -3.81 (-7.22 to -0.40) for Appendix Figure 23. Red circles indicate baseline; white circles indicate placebo; black circles indicate sulpiride.

Appendix Figure 24: Patients with chronic depression and a diagnosis of major depression or dysthymia treated with sulpiride and placebo and its effect on depressed mood[5]



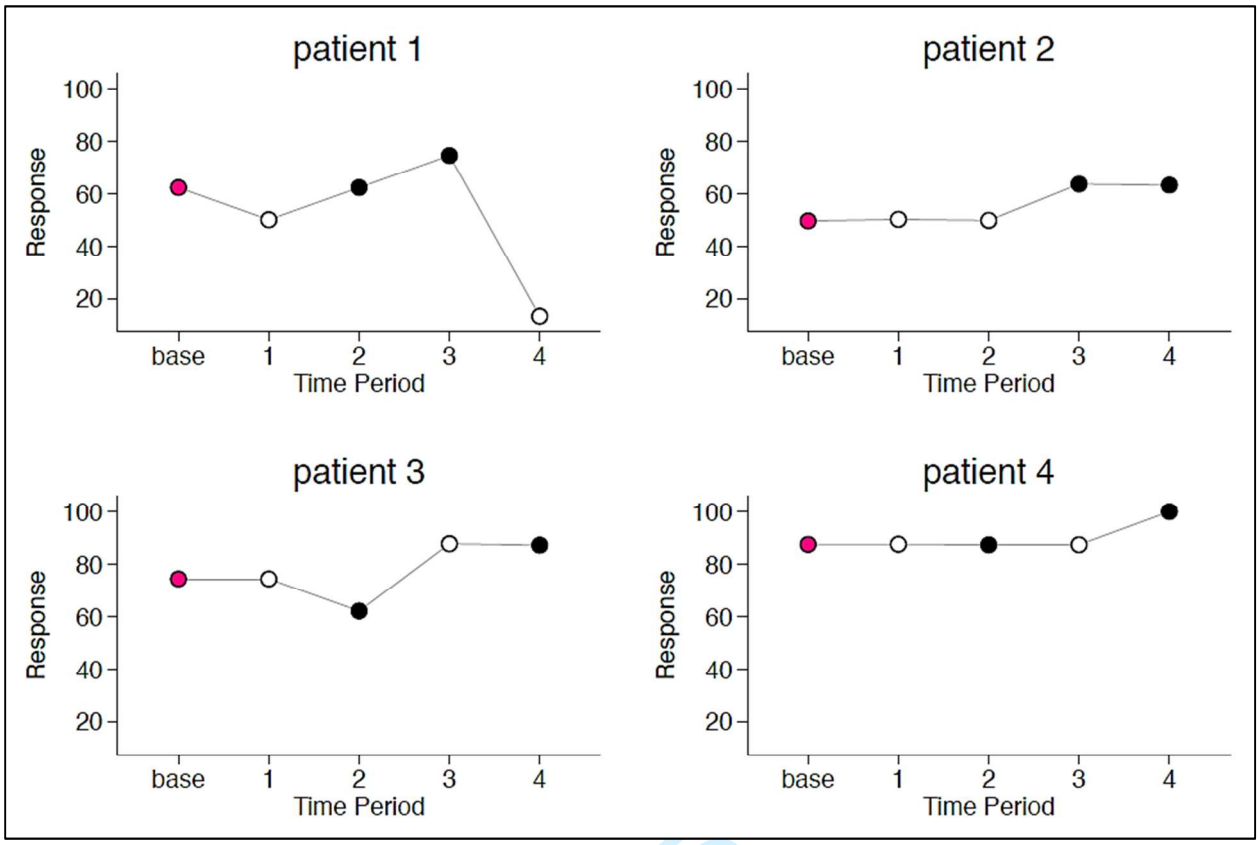
Appendix Figure 24 Legend: Data from this figure was extracted from the study published by Maier et al in 1994, which investigates the effect of sulpiride and placebo on depressed mood in patients with chronic depression and a diagnosis of major depression or dysthymia. The treatment effect is -3.63 (-7.40 to 0.15) for Appendix Figure 24. Red circles indicate baseline; white circles indicate placebo; black circles indicate sulpiride.

Appendix Figure 25: Patients with chronic depression and a diagnosis of major depression or dysthymia treated with sulpiride and placebo and its effect on somatization[5]



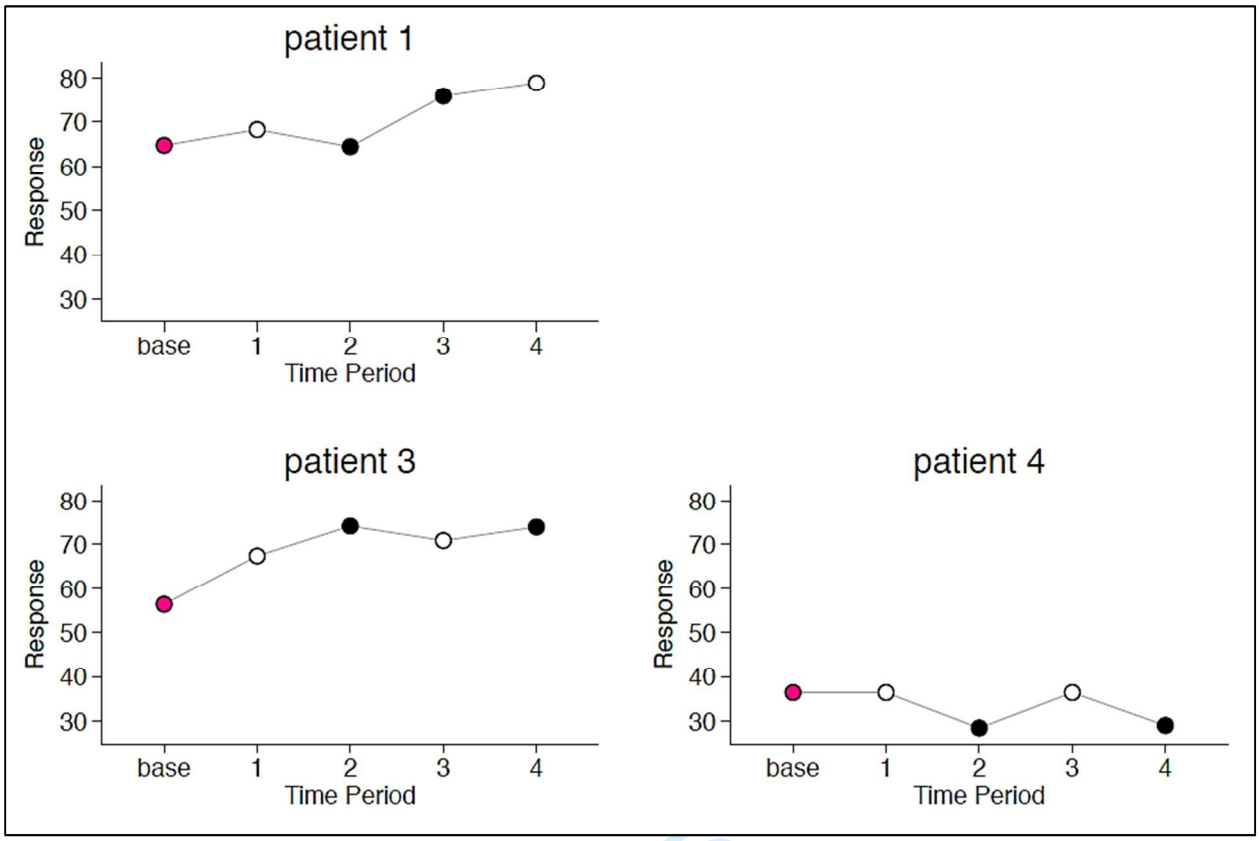
Appendix Figure 25 Legend: Data from this figure was extracted from the study published by Maier et al in 1994, which investigates the effect of sulpiride and placebo on somatization in patients with chronic depression and a diagnosis of major depression or dysthymia. The treatment effect is -1.50 (-4.20 to 1.21) for Appendix Figure 25. Red circles indicate baseline; white circles indicate placebo; black circles indicate sulpiride.

Appendix Figure 26: Patients with ataxia from traumatic brain injury treated with ondansetron and placebo and its effect on lower extremity ataxia[6]



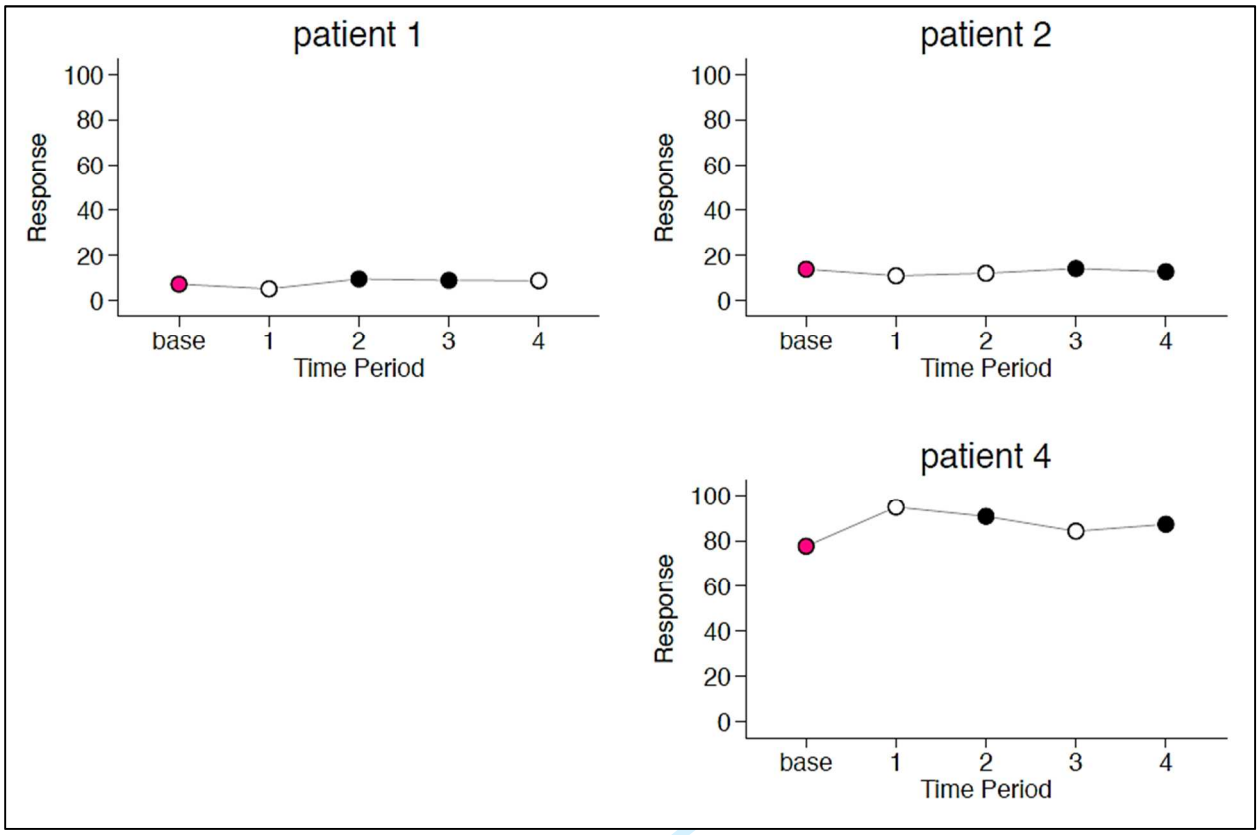
Appendix Figure 26 Legend: Data from this figure was extracted from the study published by Mandelcorn et al in 2004, which investigates the effect of ondansetron and placebo on lower extremity ataxia in patients with ataxia from traumatic brain injury. Each patient received the same treatment. The treatment effect is 12.49 (-0.85 to 25.84) for Appendix Figure 26. Red circles indicate baseline; white circles indicate placebo; black circles indicate ondansetron.

Appendix Figure 27: Patients with ataxia from traumatic brain injury treated with ondansetron and placebo and its effect on self-assessment score[6]



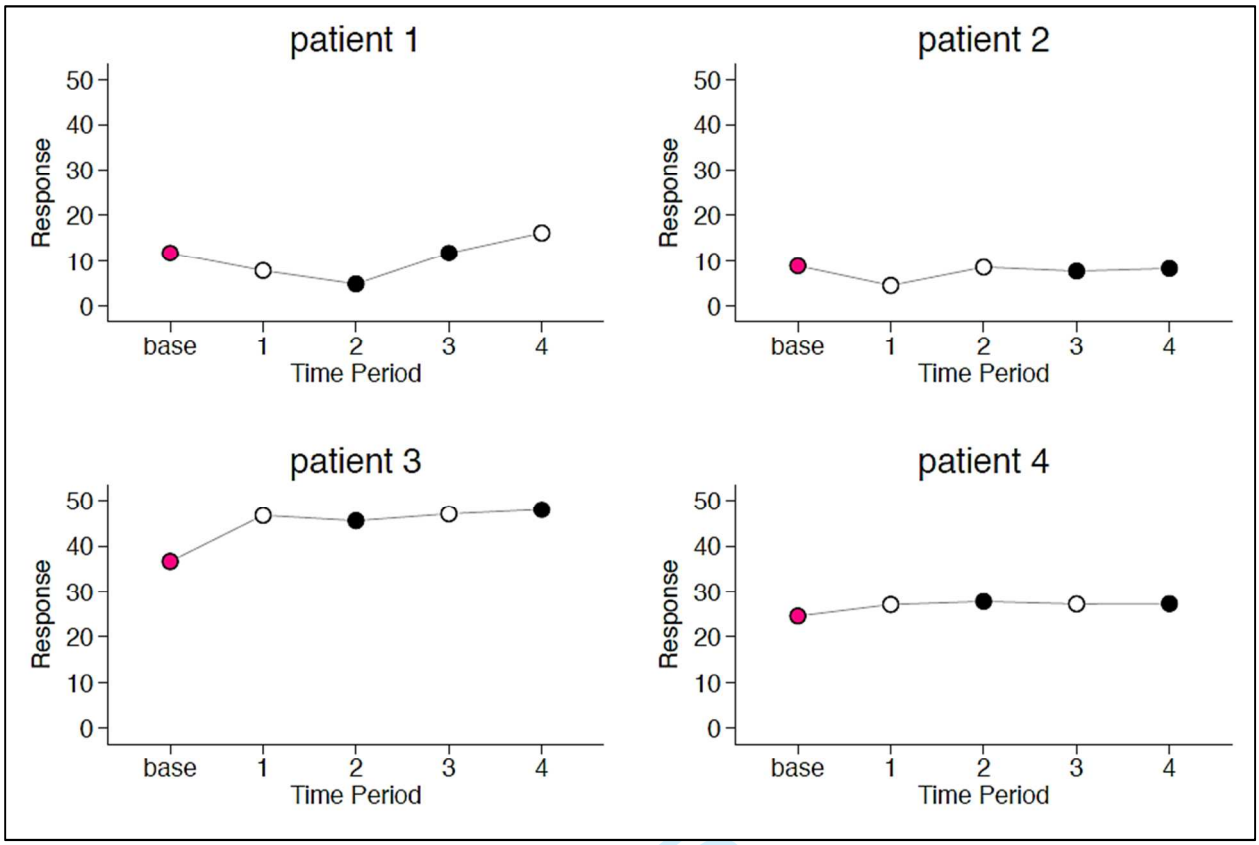
Appendix Figure 27 Legend: Data from this figure was extracted from the study published by Mandelcorn et al in 2004, which investigates the effect of ondansetron and placebo on self-assessment score in patients with ataxia from traumatic brain injury. The treatment effect is -2.05 (-8.43 to 4.33) for Appendix Figure 27. Red circles indicate baseline; white circles indicate placebo; black circles indicate ondansetron.

Appendix Figure 28: Patients with ataxia from traumatic brain injury treated with ondansetron and placebo and its effect on truncal ataxia[6]



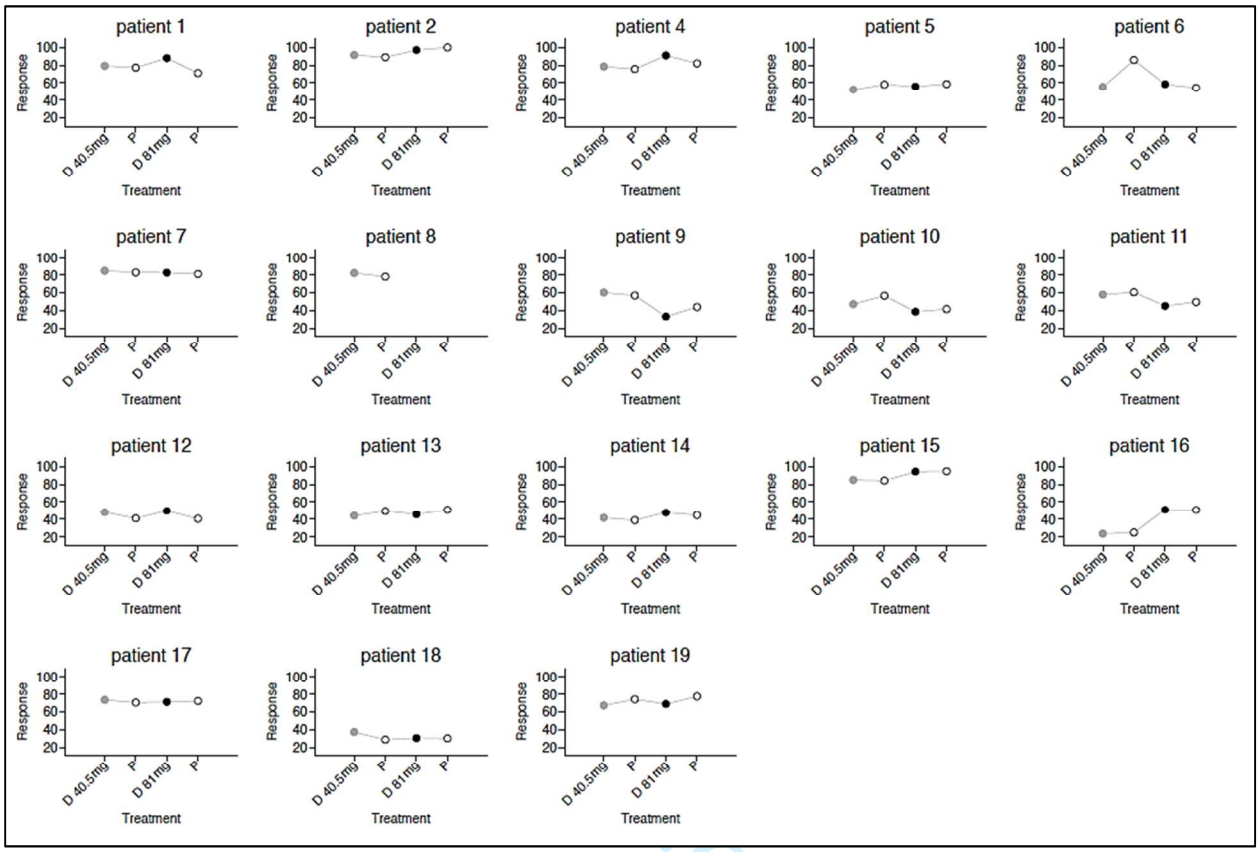
Appendix Figure 28 Legend: Data from this figure was extracted from the study published by Mandelcorn et al in 2004, which investigates the effect of ondansetron and placebo on truncal ataxia in patients with ataxia from traumatic brain injury. The treatment effect is 1.20 (-2.06 to 4.45) for Appendix Figure 28. Red circles indicate baseline; white circles indicate placebo; black circles indicate ondansetron.

Appendix Figure 29: Patients with ataxia from traumatic brain injury treated with ondansetron and placebo and its effect on upper extremity ataxia[6]



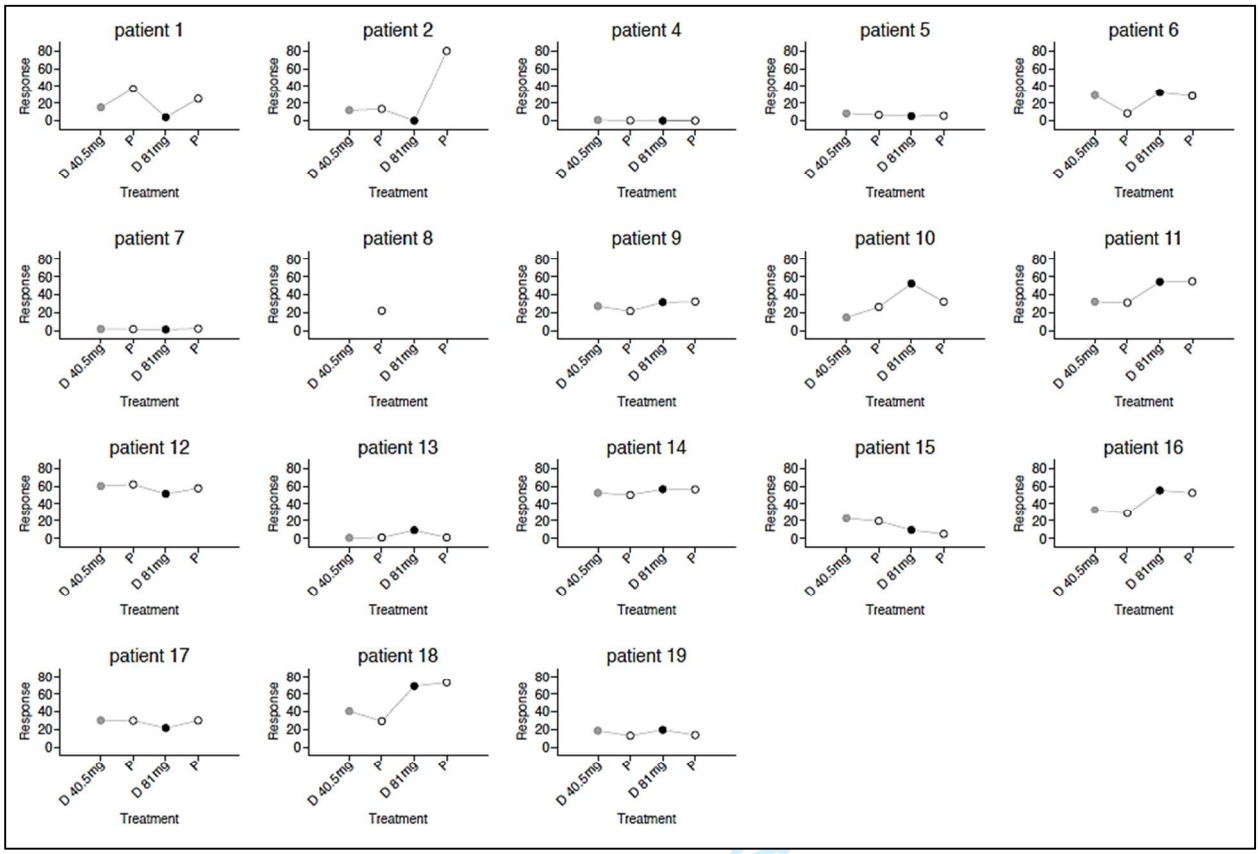
Appendix Figure 29 Legend: Data from this figure was extracted from the study published by Mandelcorn et al in 2004, which investigates the effect of ondansetron and placebo on upper extremity ataxia in patients with ataxia from traumatic brain injury. The treatment effect is -0.50 (-3.10 to 2.10) for Appendix Figure 29. Red circles indicate baseline; white circles indicate placebo; black circles indicate ondansetron.

Appendix Figure 30: Patients with chronic neuropathic pain treated with oral dextromethorphan and placebo and its effect on VAS pain intensity[7]



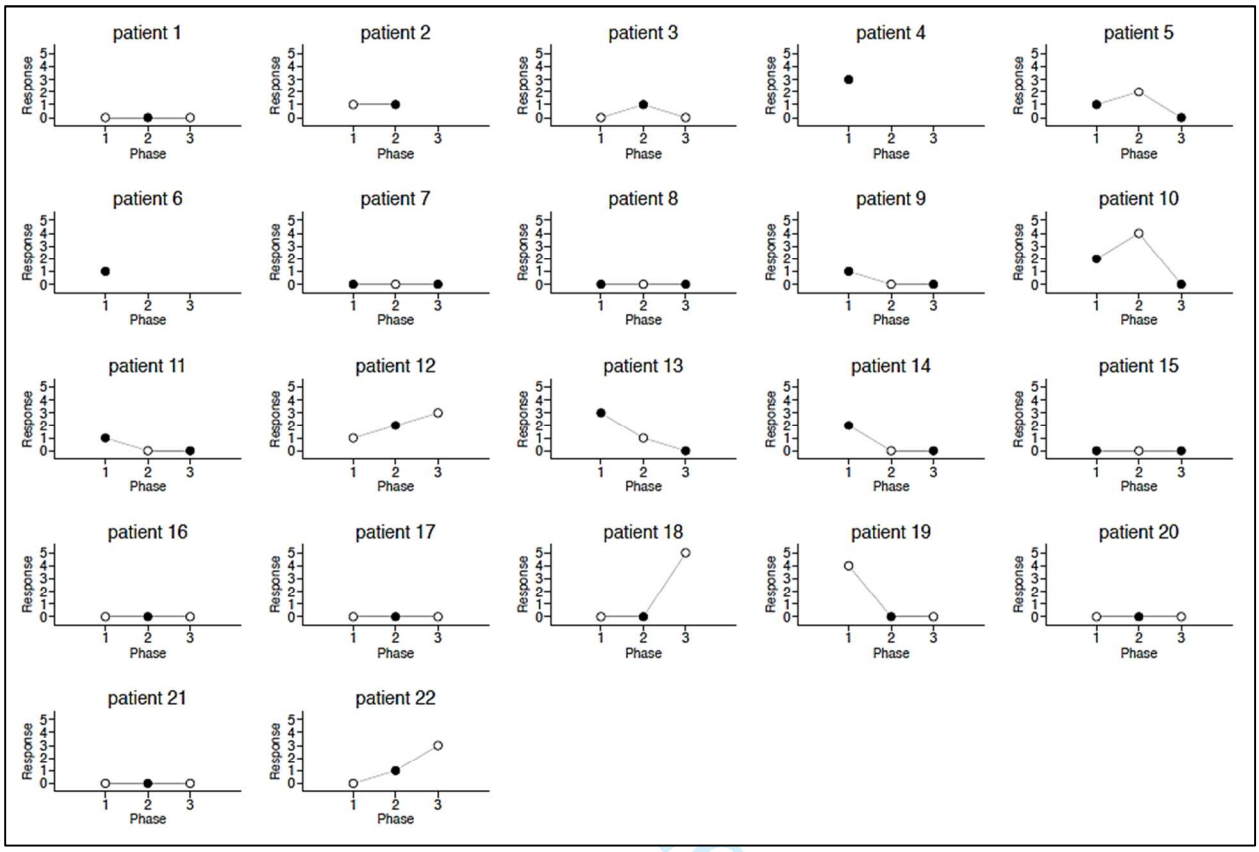
Appendix Figure 30 Legend: Data from this figure was extracted from the study published by McQuay et al in 1994, which investigates the effect of oral dextromethorphan and placebo on VAS pain intensity in patients with chronic neuropathic pain. The treatment effect is -1.06 (-5.16 to 3.04) for Appendix Figure 30. Grey circles indicate dextromethorphan 40.5 mg daily; black circles indicate dextromethorphan 81 mg daily; white circles indicate placebo.

Appendix Figure 31: Patients with chronic neuropathic pain treated with oral dextromethorphan and placebo and its effect on VAS relief intensity[7]



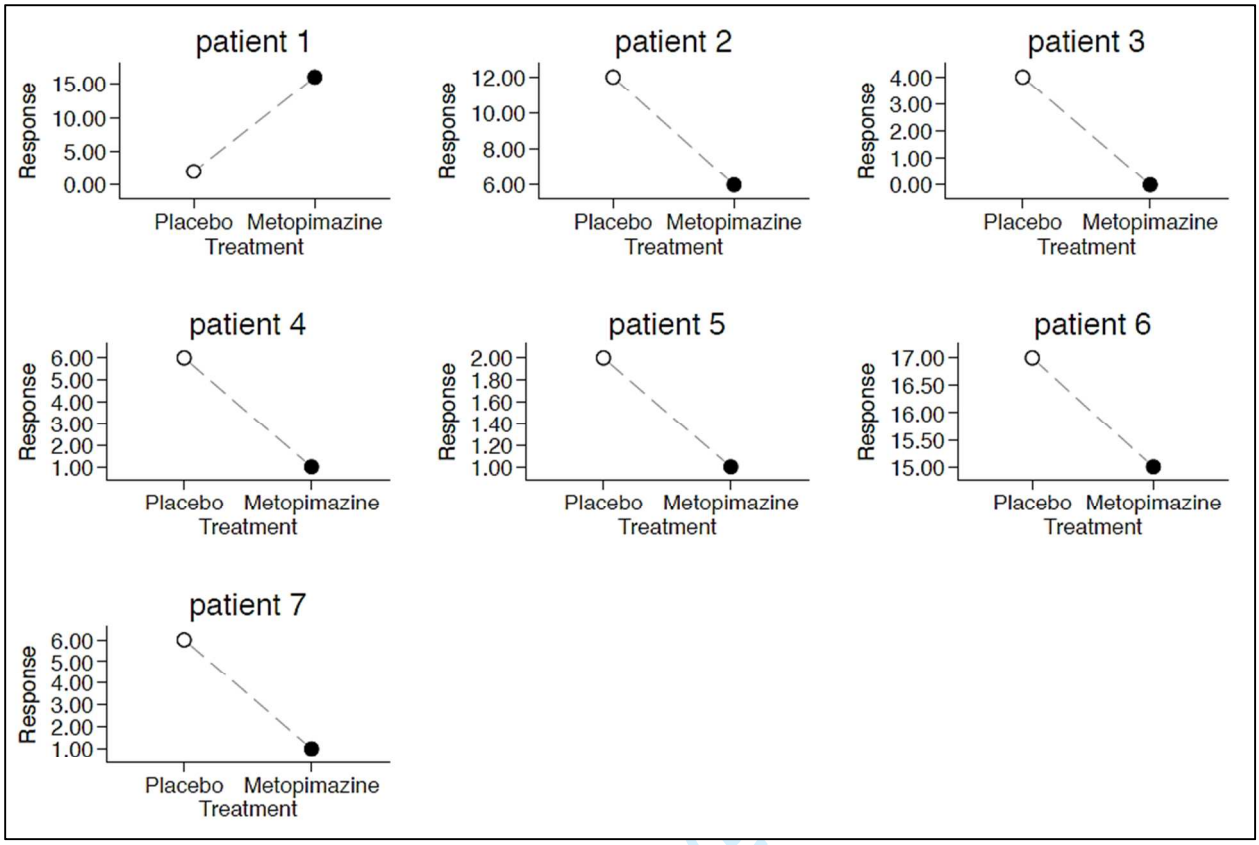
Appendix Figure 31 Legend: Data from this figure was extracted from the study published by McQuay et al in 1994, which investigates the effect of oral dextromethorphan and placebo on VAS relief intensity in patients with chronic neuropathic pain. The treatment effect is -3.86 (-11.11 to 3.40) for Appendix Figure 31. Grey circles indicate dextromethorphan 40.5 mg daily; black circles indicate dextromethorphan 81 mg daily; white circles indicate placebo.

Appendix Figure 32: Patients with unstable angina at rest treated with continuous and intermittent injection of isosorbide dinitrate and its effect on incidence of angina[8]



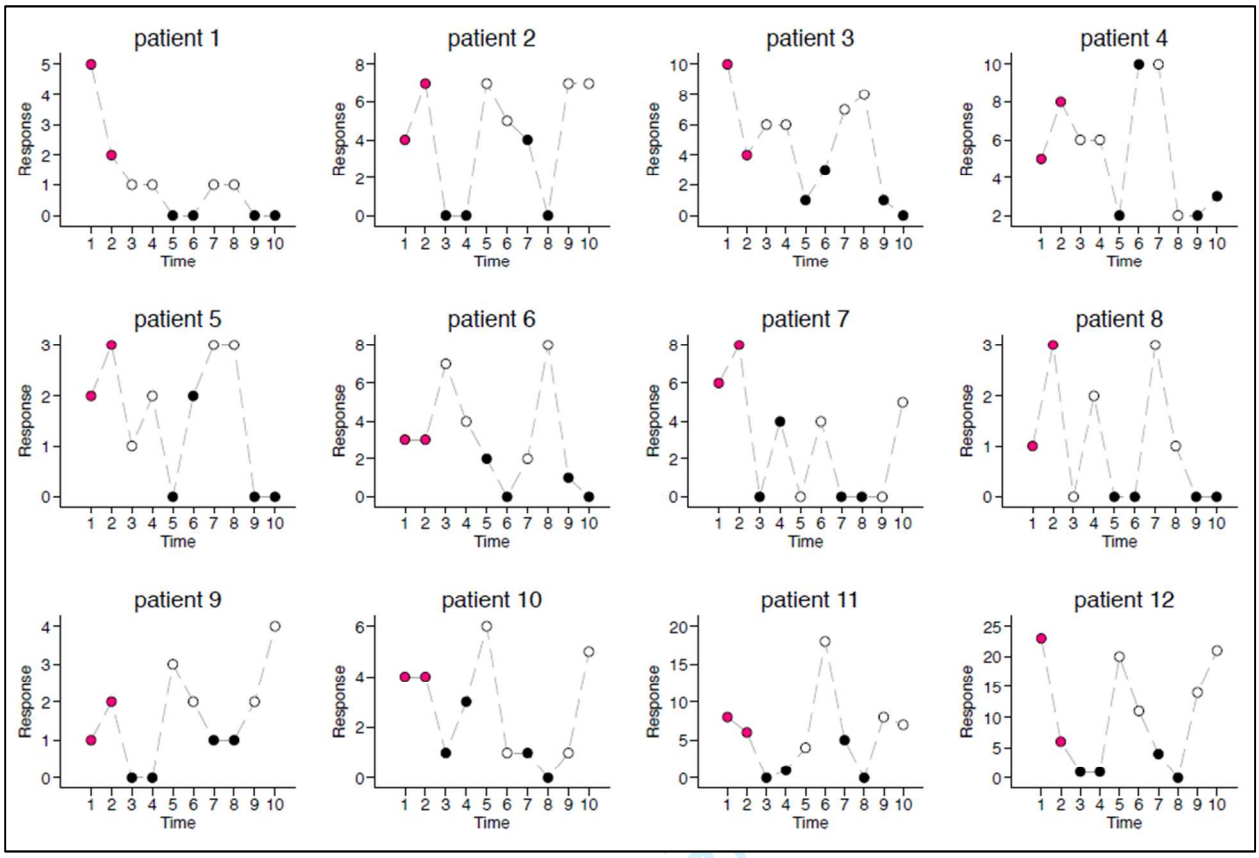
Appendix Figure 32 Legend: Data from this figure was extracted from the study published by Miyazaki et al in 1995, which investigates the effect of continuous and intermittent injection of isosorbide dinitrate on incidence of angina in patients with unstable angina. The treatment effect is 0.47 (-0.32 to 1.26) for Appendix Figure 32. White circles indicate continuous injection; black circles indicate intermittent injection.

Appendix Figure 33: Children with brain tumors receiving highly emetogenic therapy treated with ondansetron/metopimazine and ondansetron monotherapy and its effect on emetic episodes per day[9]



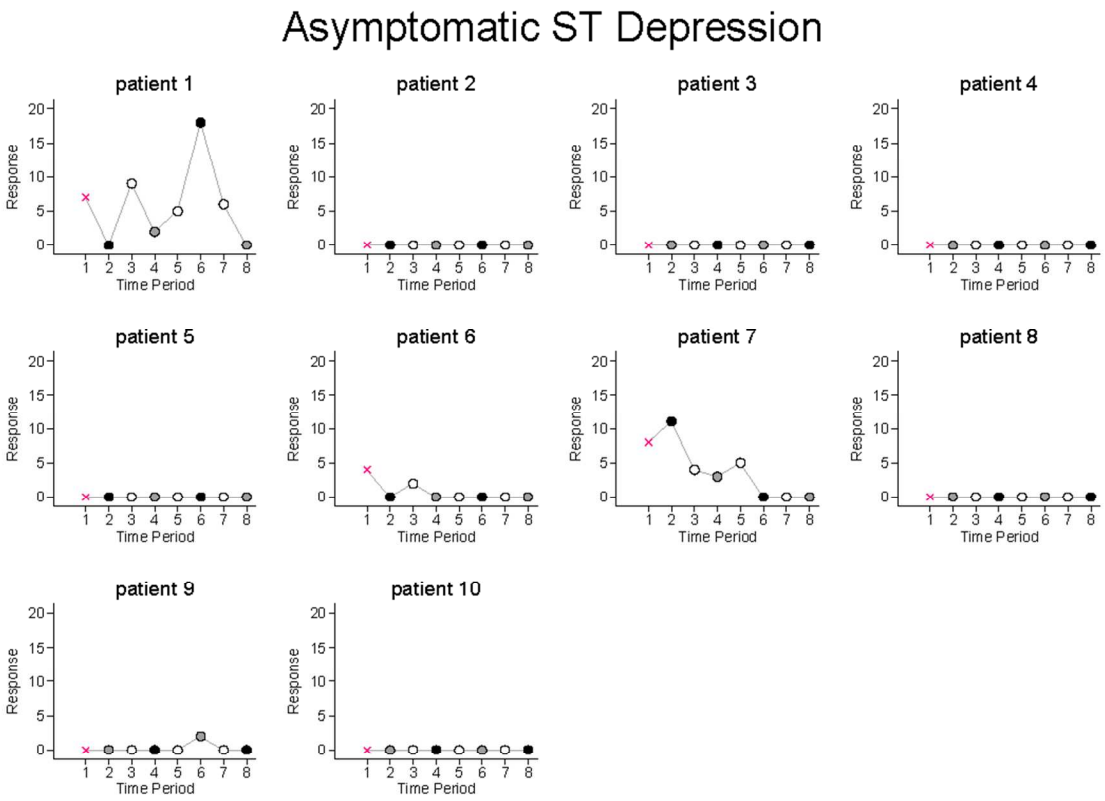
Appendix Figure 33 Legend: Data from this figure was extracted from the study published by Nathan et al in 2006, which investigates the effect of ondansetron/metopimazine and ondansetron monotherapy on emetic episodes per day in children with brain tumors receiving highly emetogenic therapy. The treatment effect is -0.56 (-1.74 to 0.62) for Appendix Figure 33. White circles indicate placebo; black circles indicate metopimazine.

Appendix Figure 34: Patients with unstable angina at rest treated with oral verapamil and placebo and its effect on ischemic attacks[10]



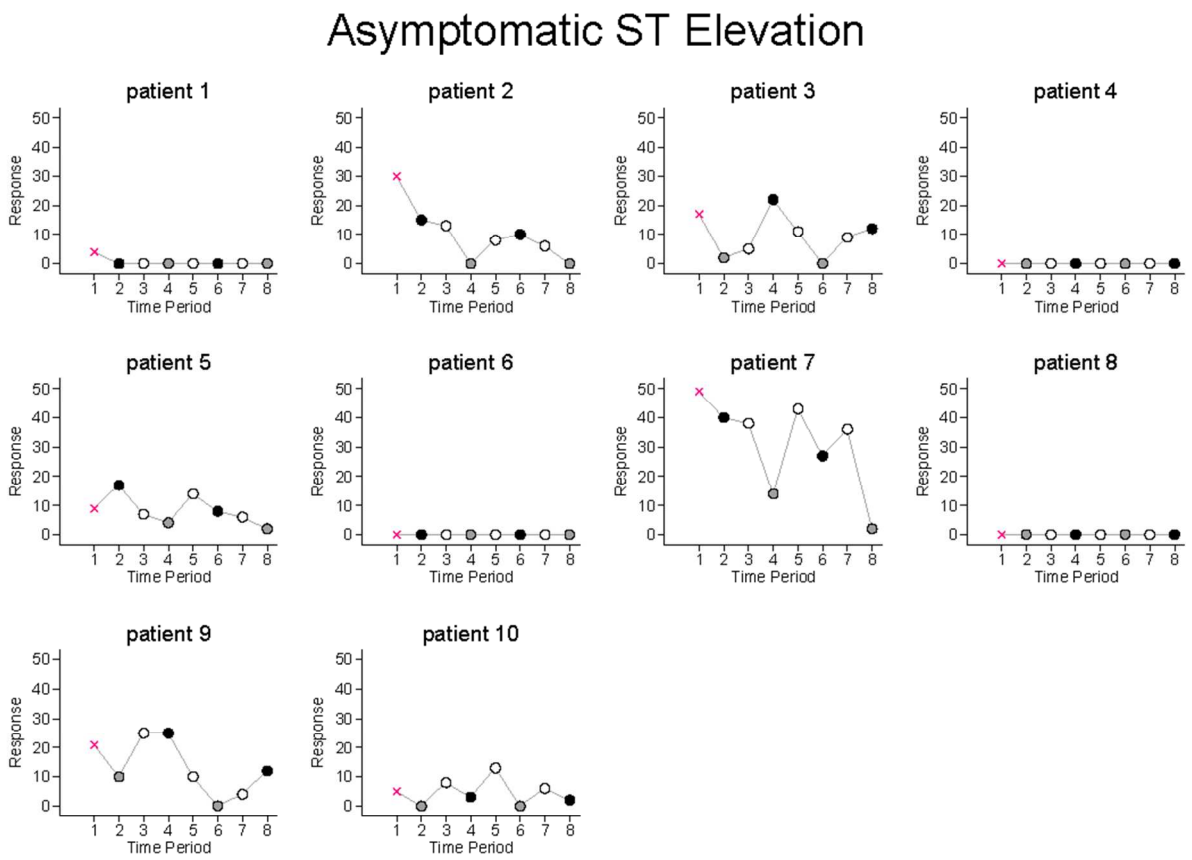
Appendix Figure 34 Legend: Data from this figure was extracted from the study published by Parodi et al in 1979, which investigates the effect of oral verapamil and placebo on ischemic attacks in patients with unstable angina. The treatment effect is -1.63 (-2.10 to -1.17) for Appendix Figure 34. Red circles indicate baseline; white circles indicate placebo; black circles indicate verapamil.

Appendix Figure 35: Patients with unstable angina at rest treated with verapamil, propranolol and placebo and its effect on asymptomatic ST depression[11]



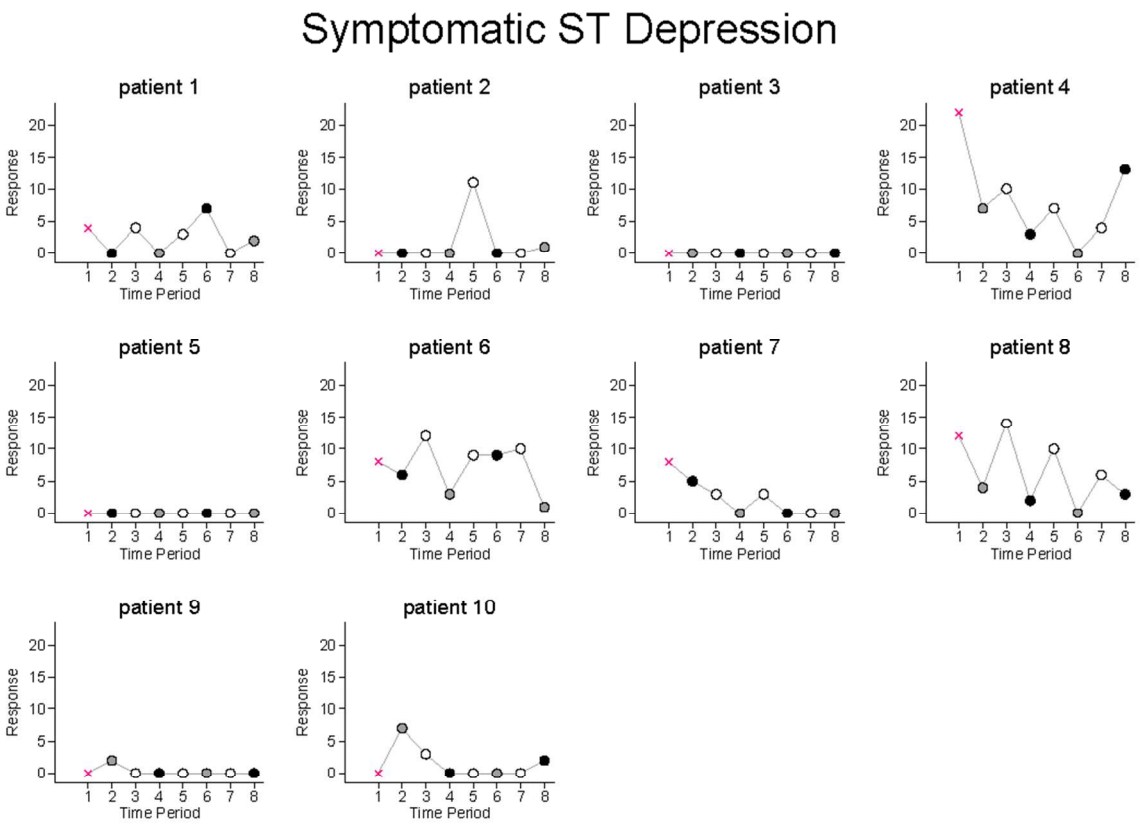
Appendix Figure 35 Legend: Data from this figure was extracted from the study published by Parodi et al in 1986, which investigates the effect of verapamil, propranolol and placebo on asymptomatic ST depression in patients with unstable angina. The treatment effect is -0.82 (-2.54 to 0.90) for Appendix Figure 35. Red Xs indicate baseline; white circles indicate placebo; grey circles indicate propranolol; black circles indicate verapamil.

Appendix Figure 36: Patients with unstable angina at rest treated with verapamil, propranolol and placebo and its effect on asymptomatic ST elevation[11]



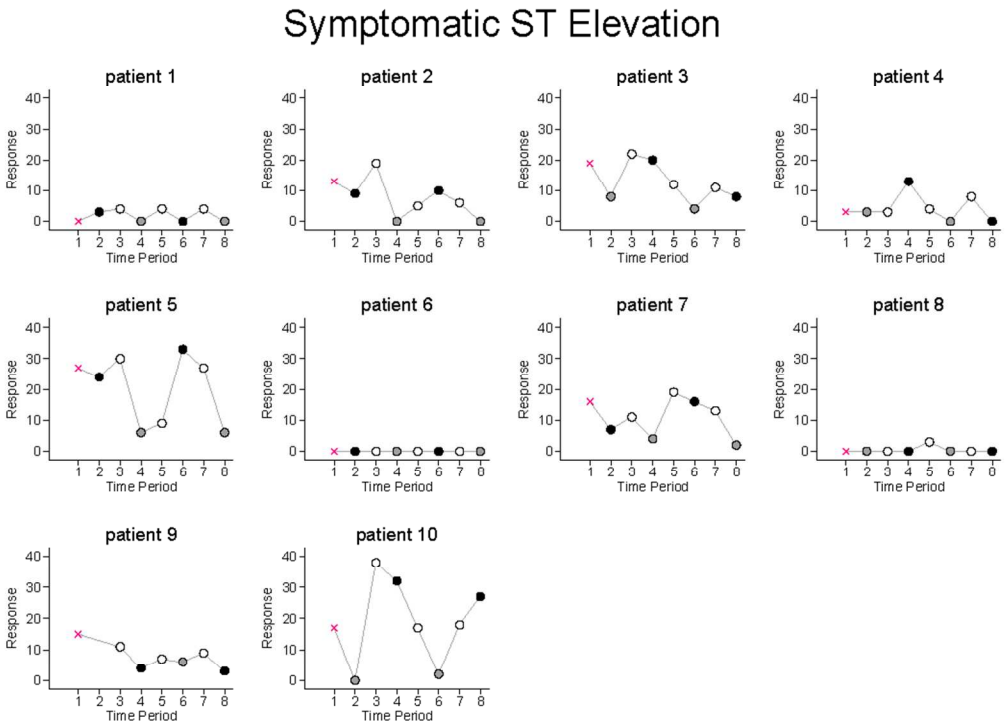
Appendix Figure 36 Legend: Data from this figure was extracted from the study published by Parodi et al in 1986, which investigates the effect of verapamil, propranolol and placebo on asymptomatic ST elevation in patients with unstable angina. The treatment effect is -1.97 (-2.92 to -1.01) for Appendix Figure 36. Red Xs indicate baseline; white circles indicate placebo; grey circles indicate propranolol; black circles indicate verapamil.

Appendix Figure 37: Patients with unstable angina at rest treated with verapamil, propranolol and placebo and its effect on symptomatic ST depression[11]



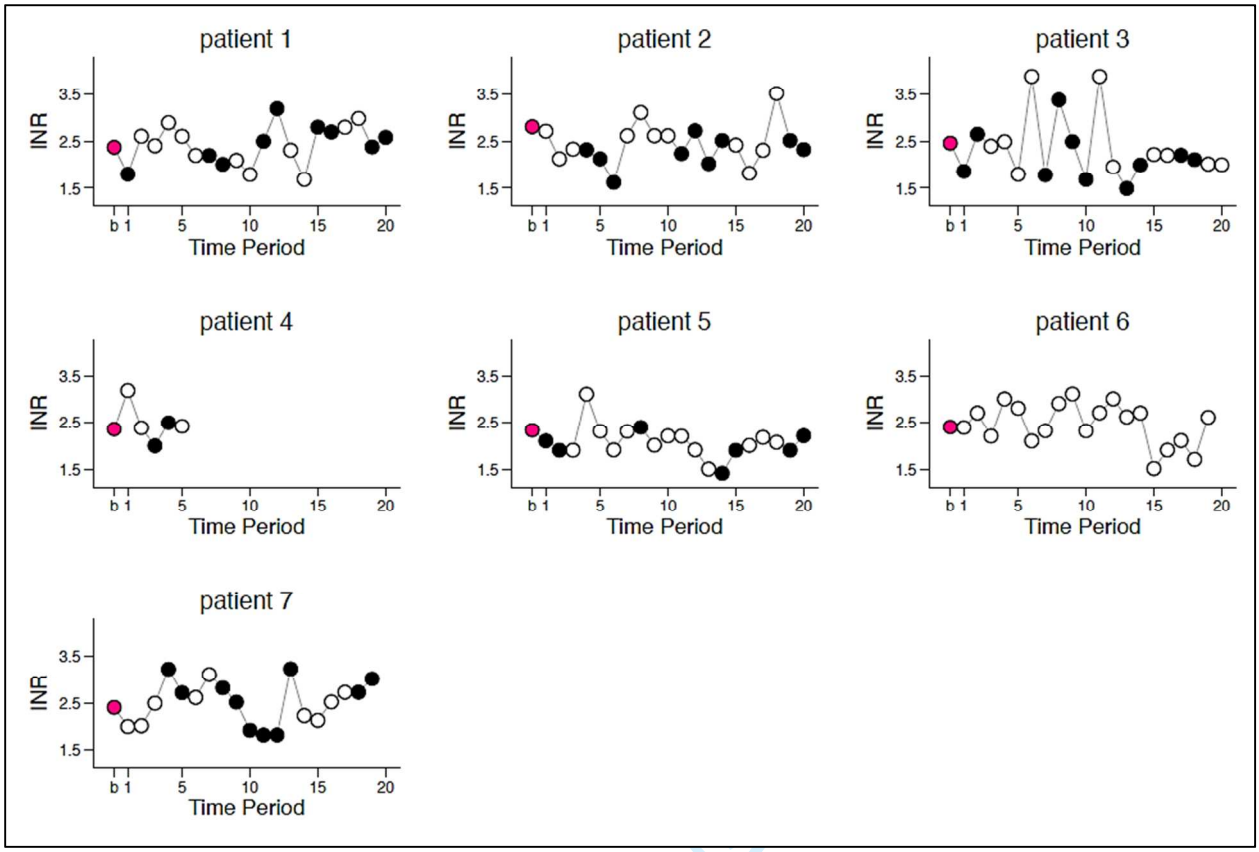
Appendix Figure 37 Legend: Data from this figure was extracted from the study published by Parodi et al in 1986, which investigates the effect of verapamil, propranolol and placebo on symptomatic ST depression in patients with unstable angina. The treatment effect is -0.98 (-1.84 to -0.13) for Appendix Figure 37. Red Xs indicate baseline; white circles indicate placebo; grey circles indicate propranolol; black circles indicate verapamil.

Appendix Figure 38: Patients with unstable angina at rest treated with verapamil, propranolol and placebo and its effect on symptomatic ST elevation[11]



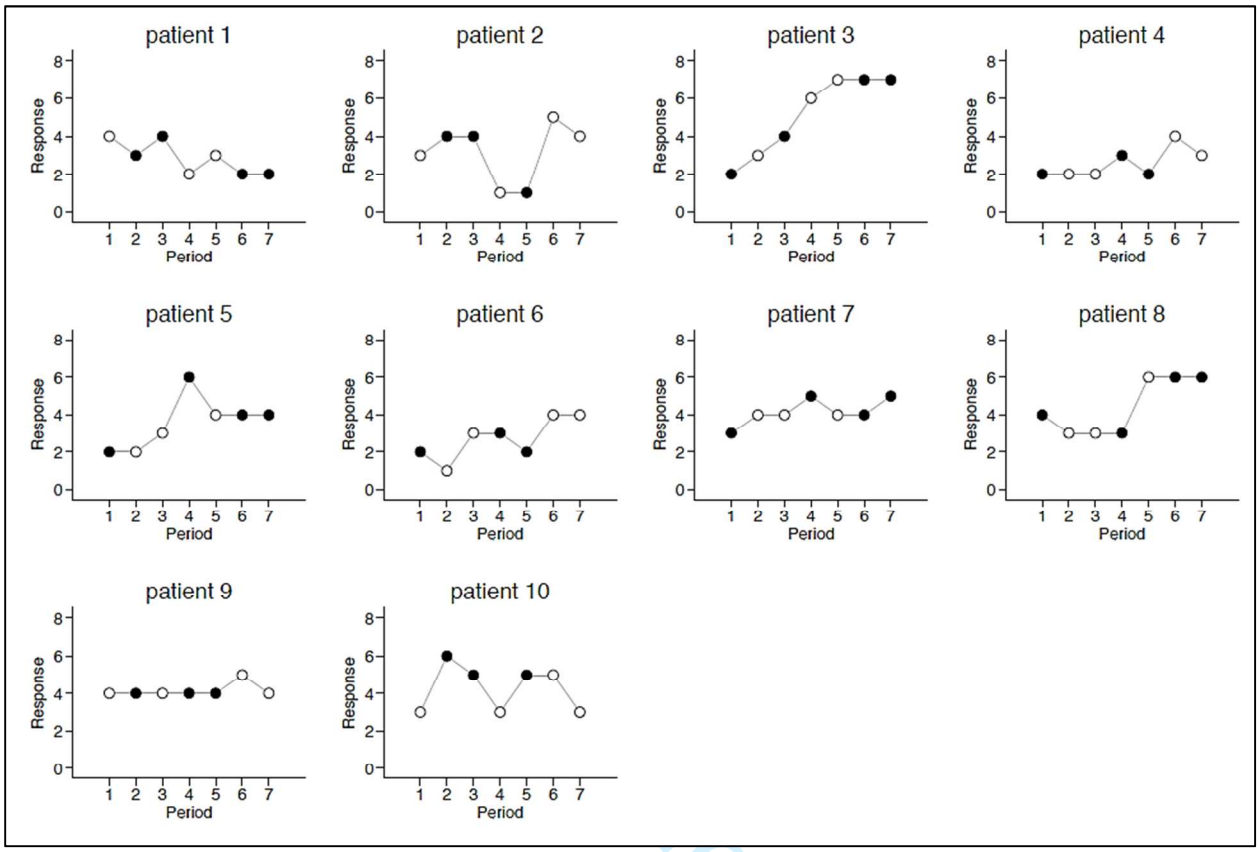
Appendix Figure 38 Legend: Data from this figure was extracted from the study published by Parodi et al in 1986, which investigates the effect of verapamil, propranolol and placebo on symptomatic ST elevation in patients with unstable angina. The treatment effect is -1.87 (-2.72 to -1.02) for Appendix Figure 38. Red Xs indicate baseline; white circles indicate placebo; grey circles indicate propranolol; black circles indicate verapamil.

Appendix Figure 39: Patients previously taking warfarin for either atrial fibrillation or deep vein thrombosis treated with apo-warfarin and coumadin and its effect on international normalized ratio[12]



Appendix Figure 39 Legend: Data from this figure was extracted from the study published by Pereira et al in 1995, which investigates the effect of apo-warfarin and coumadin on international normalized ratio in patients previously taking warfarin for either atrial fibrillation or deep vein thrombosis. The treatment effect is -0.12 (-0.30 to 0.07) for Appendix Figure 39. Red circles indicate baseline; white circles indicate Coumadin; black circles indicate apo-warfarin.

Appendix Figure 40: Parkinson's disease patients with troublesome dyskinesia treated with simvastatin and placebo and its effect on discomfort caused by troublesome dyskinesia[13]



Appendix Figure 40 Legend: Data from this figure was extracted from the study published by Tison et al in 2012, which investigates the effect of simvastatin and placebo on discomfort caused by troublesome dyskinesia in Parkinson's disease patients with troublesome dyskinesia. The treatment effect is0.20 (-0.40 to 0.80) for Appendix Figure 40. White circles indicate placebo; black circles indicate simvastatin.

Appendix Reference List

(1) Camfield P, Gordon K, Dooley J, Camfield C. Melatonin appears ineffective in children with intellectual deficits and fragmented sleep: six "N of 1" trials. *J Child Neurol* 1996; 11(4):341-343.

(2) Hinderer SR. The supraspinal anxiolytic effect of baclofen for spasticity reduction. *Am J Phys Med Rehabil* 1990; 69(5):254-258.

(3) Langer JC, Winthrop AL, Issenman RM. The single-subject randomized trial. A useful clinical tool for assessing therapeutic efficacy in pediatric practice. *Clin Pediatr (Phila)* 1993; 32(11):654-657.

(4) Lashner BA, Hanauer SB, Silverstein MD. Testing nicotine gum for ulcerative colitis patients. Experience with single-patient trials. *Dig Dis Sci* 1990; 35(7):827-832.

(5) Maier W, Benkert O. Treatment of chronic depression with sulpiride: evidence of efficacy in placebo-controlled single case studies. *Psychopharmacology (Berl)* 1994; 115(4):495-501.

(6) Mandelcorn J, Cullen NK, Bayley MT. A preliminary study of the efficacy of ondansetron in the treatment of ataxia, poor balance and incoordination from brain injury. *Brain Inj* 2004; 18(10):1025-1039.

(7) McQuay HJ, Carroll D, Jadad AR, Glynn CJ, Jack T, Moore RA et al. Dextromethorphan for the treatment of neuropathic pain: a double-blind randomised controlled crossover trial with integral n-of-1 design. *Pain* 1994; 59(1):127-133.

(8) Miyazaki S, Nonogi H, Goto Y, Sumiyoshi T, Haze K, Hiramori K. Comparison of the therapeutic efficacy of continuous and intermittent injection of isosorbide dinitrate: a randomized study on unstable angina. *Intern Med* 1995; 34(9):856-862.

(9) Nathan PC, Tomlinson G, Dupuis LL, Greenberg ML, Ota S, Bartels U et al. A pilot study of ondansetron plus metopimazine vs. ondansetron monotherapy in children receiving highly emetogenic chemotherapy: a Bayesian randomized serial N-of-1 trials design. *Support Care Cancer* 2006; 14(3):268-276.

(10) Parodi O, Maseri A, Simonetti I. Management of unstable angina at rest by verapamil. A double-blind cross-over study in coronary care unit. *Br Heart J* 1979; 41(2):167-174.

(11) Parodi O, Simonetti I, Michelassi C, Carpeggiani C, Biagini A, L'Abbate A et al. Comparison of verapamil and propranolol therapy for angina pectoris at rest: a randomized, multiple-crossover, controlled trial in the coronary care unit. *Am J Cardiol* 1986; 57(11):899-906.

(12) Pereira JA, Holbrook AM, Dolovich L, Goldsmith C, Thabane L, Douketis JD et al. Are brand-name and generic warfarin interchangeable? Multiple n-of-1 randomized, crossover trials. *Ann Pharmacother* 2005; 39(7-8):1188-1193.

(13) Tison F, Negre-Pages L, Meissner WG, Dupouy S, Li Q, Thiolat ML et al. Simvastatin decreases levodopa-induced dyskinesia in monkeys, but not in a randomized, placebo-controlled, multiple

cross-over ("n-of-1") exploratory trial of simvastatin against levodopa-induced dyskinesia in Parkinson's disease patients. *Parkinsonism Relat Disord* 2013; 19(4):416-421.

(14) Emmanuel AV, Kamm MA, Roy AJ, Kerstens R, Vandeplassche L. Randomised clinical trial: the efficacy of prucalopride in patients with chronic intestinal pseudo-obstruction--a double-blind, placebo-controlled, cross-over, multiple n = 1 study. *Aliment Pharmacol Ther* 2012; 35(1):48-55.

(15) Haas DC, Sheehe PR. Dextroamphetamine pilot crossover trials and n of 1 trials in patients with chronic tension-type and migraine headache. *Headache* 2004; 44(10):1029-1037.

(16) Jaeschke R, Adachi J, Guyatt G, Keller J, Wong B. Clinical usefulness of amitriptyline in fibromyalgia: the results of 23 N-of-1 randomized controlled trials. *J Rheumatol* 1991; 18(3):447-451.

(17) Johannessen T, Petersen H, Kristensen P, Fosstvedt D, Kleveland PM, Dybdahl J et al. Cimetidine on-demand in dyspepsia. Experience with randomized controlled single-subject trials. *Scand J Gastroenterol* 1992; 27(3):189-195.

(18) Mahon J, Laupacis A, Donner A, Wood T. Randomised study of n of 1 trials versus standard practice. *BMJ* 1996; 312(7038):1069-1074.

(19) March L, Irwig L, Schwarz J, Simpson J, Chock C, Brooks P. n of 1 trials comparing a non-steroidal anti-inflammatory drug with paracetamol in osteoarthritis. *BMJ* 1994; 309(6961):1041-1045.

(20) Patel A, Jaeschke R, Guyatt GH, Keller JL, Newhouse MT. Clinical usefulness of n-of-1 randomized controlled trials in patients with nonreversible chronic airflow limitation. *Am Rev Respir Dis* 1991; 144(4):962-964.

(21) Wallace AE, Kofoed LL. Statistical analysis of single case studies in the clinical setting: the example of methylphenidate trials in children with attention-deficit hyperactivity disorder. *J Child Adolesc Psychopharmacol* 1994; 4(3):141-150.

(22) Woodfield R, Goodyear-Smith F, Arroll B. N-of-1 trials of quinine efficacy in skeletal muscle cramps of the leg. *Br J Gen Pract* 2005; 55(512):181-185.

(23) Zucker DR, Ruthazer R, Schmid CH, Feuer JM, Fischer PA, Kieval RI et al. Lessons learned combining N-of-1 trials to assess fibromyalgia therapies. *J Rheumatol* 2006; 33(10):2069-2077.

Appendix Table 4. Studies reporting person-level treatment effect with both fixed-effect and random-effect using a method of moments estimator

Study	Outcome	Fixed effect model	P for HTE (fixed-effects model)	Random Treatment Effect	summary_tau2	P for HTE (random-effects model)	I-square
March 1999	Mean pain score on VAS taken from 2nd week of tx	-4.155 (-4.807 to -3.502)	<0.001	-7.093 (-11.939 to -2.248)	73.530	<0.001	97.5%
March 1999	Mean stiffness score on VAS taken from 2nd week of	-2.192 (-2.549 to -1.835)	<0.001	-5.992 (-11.280 to -0.704)	88.872	<0.001	97.5%
Emmanuel 2012	Bloating	-0.131 (-0.171 to -0.090)	<0.001	-0.344 (-0.619 to -0.069)	0.071	<0.001	94.2%
Emmanuel 2012	Pain	-0.160 (-0.209 to -0.111)	<0.001	-0.440 (-0.771 to -0.110)	0.106	<0.001	96.0%
Haas 2004	Chronic tension-type headache grade	0.733 (0.609 to 0.857)	<0.001	0.772 (0.454 to 1.090)	0.350	<0.001	84.4%
Haas_2004	Chronic tension-type headache grade	0.543 (0.394 to 0.693)	0.067	0.542 (0.354 to 0.731)	0.055	0.067	37.2%
Jaeschke 1991	7-point symptom scale	0.356 (0.286 to 0.426)	<0.001	0.427 (0.210 to 0.645)	0.186	<0.001	85.0%
Jaeschke 1991	Tender point changes count	1.072 (0.701 to 1.443)	<0.001	1.320 (0.404 to 2.236)	2.166	<0.001	72.3%
Johannessen 1992	6-point symptom scale	0.657 (0.530 to	<0.001	0.698 (0.466 to 0.931)	0.382	<0.001	65.8%

		0.785)					
Mahon 1996	Likert Scale (1-7)	0.069 (-0.042 to 0.179)	<0.001	0.145 (-0.153 to 0.443)	0.134	<0.001	77.6%
Patel 1991	4-item symptom questionnaire	0.000 (-0.000 to 0.000)	<0.001	0.000 (-0.000 to 0.000)	0.000	<0.001	90.9%
Pereira 1995	INR (diff)	0.027 (-0.155 to 0.209)	0.477	0.027 (-0.155 to 0.209)	0.000	0.477	0.0%
Wallace 1994		0.759 (0.341 to 1.178)	0.747	0.759 (0.341 to 1.178)	0.000	0.747	0.0%
Woodfield 15808032	Number of cramps	-5.395 (-7.091 to -3.699)	<0.001	-18.823 (-28.527 to -9.120)	161.582	<0.001	92.0%
Woodfield 15808032	Total days with cramps	-7.600 (-8.420 to -6.781)	<0.001	-6.181 (-9.798 to -2.563)	26.245	<0.001	93.6%
Zucker 2006	FIQ	-5.019 (-8.784 to -1.254)	0.999	-5.019 (-8.784 to -1.254)	0.000	0.999	0.0%

Appendix Table 5. Studies reporting person-level outcomes with both fixed-effect and random-effect hierarchical linear model

Author Year	Outcome	Range of the Scales (severity)	Fixed Treatment Effect	Random Treatment Effect	P-value Person Treatment Interaction
Camfield 1996	Nights without awakening	NR	0.865 (0.215 to 1.516)	0.84 (0.20 to 1.48)	0.456
Hinderer 1990	Anxiety		0.000 (0.000 to 0.000)	-1.06 (-1.88 to -0.23)	<0.001
Langer 1993	Vomiting	NR	-1.204 (-2.494 to 0.086)	-1.20 (-2.49 to 0.09)	0.136
Lashner 1990	Symptom score: abdominal pain	Symptom scores 0-100 (0=best, 100=worst)	-3.615 (-16.982 to 9.751)	-3.62 (-15.84 to 8.61)	0.007
	Symptom score: bowel movements/day		-0.538 (-1.215 to 0.138)	-0.56 (-1.22 to 0.09)	0.001
	Symptom score: consistency of bowel movements		7.000 (-7.551 to 21.551)	7.00 (-6.29 to 20.29)	0.013
	Symptom score: hematochezia		2.308 (-17.210 to 21.826)	2.35 (-17.21 to 21.90)	0.003
	Symptom score: general sense of well-being		-6.538 (-25.352 to 12.275)	-6.54 (-23.62 to 10.56)	0.008
Maier 1994	SCL-90 subscales: Depressed mood	NR	-3.536 (-6.718 to -0.354)	-3.63 (-7.40 to 0.15)	<0.001
	SCL-90 subscales: Anxiety		-3.753 (-6.582 to -0.924)	-3.81 (-7.22 to -0.40)	<0.001
	SCL-90 subscales: Somatization		-1.419 (-4.316 to 1.478)	-1.50 (-4.20 to 1.21)	0.869
Mandelcorn 2004	Self-Assessment score	0–5 (0=worst, 5=best)	-2.052 (-8.865 to 4.761)	-2.05 (-8.43 to 4.33)	0.05
	Lower extremity ataxia	Fugl-Meyer: 3-point (0 cannot be performed to 2 can be fully performed)	12.494 (-3.155 to 28.142)	12.49 (-0.85 to 25.84)	0.025
	Truncal ataxia	AMTI forceplate®: NR	1.196 (-2.866 to 5.257)	1.20 (-2.06 to 4.45)	0.690

Author Year	Outcome	Range of the Scales (severity)	Fixed Treatment Effect	Random Treatment Effect	P-value Person Treatment Interaction
		Berg Balance Scale® 0–56, with a higher score indicating a better performance			
	Upper extremity ataxia	Purdue Pegboard Test®: pegs inserted into the board with each hand in 30 sec *Minnesota Placing Test®: reach out, grasp, and place blocks in a specific order*	-0.498 (-3.546 to 2.550)	-0.50 (-3.10 to 2.10)	0.382
McQuay 1994	VAS Pain Intensity	0-100 (0 = no pain, 100 = worst possible pain)	-1.094 (-5.572 to 3.383)	-1.06 (-5.16 to 3.04)	0.004
	VAS Relief Intensity	0-100 (0 = no relief, 100 =complete pain relief)	-3.913 (-11.729 to 3.903)	-3.86 (-11.11 to 3.40)	0.038
Miyazaki 1995	Incidence of angina	Either ST-segment elevation or depression at rest	0.496 (-0.206 to 1.199)	0.47 (-0.32 to 1.26)	0.125
Nathan 2006	Emetic episodes per day	complete response (0 episodes/day), major response (1–2 episodes/day), or failure (>2 episodes/day)	-0.095 (-0.514 to 0.325)	-0.56 (-1.74 to 0.62)	0.001
Parodi 1979	Ischemic attacks	ST elevation or depression (details NR)	-1.544 (-1.838 to -1.251)	-1.63 (-2.10 to -1.17)	0.007
Parodi 1986	Asymptomatic ST elevation (After verapamil)	NR	-1.637 (-1.994 to -1.279)	-1.97 (-2.92 to -1.01)	0.110
	Asymptomatic ST depression (After verapamil)		-1.083 (-1.903 to -0.262)	-0.82 (-2.54 to 0.90)	0.401

Author Year	Outcome	Range of the Scales (severity)	Fixed Treatment Effect	Random Treatment Effect	P-value Person Treatment Interaction
	Symptomatic ST elevation (After verapamil)		-1.580 (-1.906 to -1.254)	-1.87 (-2.72 to -1.02)	<0.001
	Symptomatic ST Depression (After verapamil)		-0.990 (-1.411 to -0.569)	-0.98 (-1.84 to -0.13)	0.002
	Asymptomatic ST elevation (After propranolol)		0.100 (-0.086 to 0.286)	-1.966 (-2.917 to -1.014)	0.006
	Asymptomatic ST depression (After propranolol)		0.339 (-0.168 to 0.845)	-0.821 (-2.539 to 0.897)	0.964
	Symptomatic ST elevation (After propranolol)		-0.002 (-0.177 to 0.173)	-1.868 (-2.718 to -1.017)	0.063
	Symptomatic ST Depression (After propranolol)		-0.374 (-0.709 to -0.039)	-0.981 (-1.835 to -0.126)	0.023
Pereira 1995	INR	Target INR range of 2.0–3.0		-0.12 (-0.30 to 0.07)	0.433
Tison 2012	Troublesome dyskinesia	7 points scale (1=extremely uncomfortable, 7=not at all uncomfortable)		0.20 (-0.40 to 0.80)	0.593

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	1-2
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	2
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	n/a
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5-6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5-6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	a1-a3
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	6-7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	6-8
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	n/a
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	8-9
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	8-9

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	n/a
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	8-9
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	9, 20, 21, 29
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	1-12, 22-26
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	n/a
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	31, a11-a50
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	10-12, 26
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	n/a
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	12, a53-a57
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	12-14
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	15-16
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	16
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	17

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

PRISMA 2009 Checklist

BMJ Open

Evaluation of person-level heterogeneity of treatment effects in published multi-person N-of-1 studies: systematic review and re-analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-017641.R1
Article Type:	Research
Date Submitted by the Author:	01-Dec-2017
Complete List of Authors:	Raman, G; Tufts Medical Center Balk, EM; Brown University Lai, Lana; Tufts Medical Center Shi, Jennifer; Tufts Medical Center Chan, Jeffrey; VA Boston Healthcare System, Center for Healthcare Organization and Implementation Research (CHOIR) Lutz, Jennifer; Tufts Medical Center Dubois, Robert; National Pharmaceutical Council, Research Kravitz, Richard; University of California Davis Kent, David; Tufts Medical Center
Primary Subject Heading:	Patient-centred medicine
Secondary Subject Heading:	Research methods
Keywords:	perseonalized medicine, n-of-1 studies, systematic review, heterogeneity of treatment effect

Evaluation of person-level heterogeneity of treatment effects in published multi-person N-of-1 studies: systematic review and re-analysis

Gowri Raman, MD, MS[a]; Ethan M Balk, MD, MPH[b]; Lana Lai, MS[c]; Jennifer Shi, BA[d]; Jeff Chan, MD[a,e]; Jennifer Lutz, MA[c]; Robert Dubois, MD, PhD[f]; Richard L Kravitz, MD, MSPH[g]; David M Kent, MD, MS*[c]

[a]Center for Clinical Evidence Synthesis, ICRHPS, Medical Center/Tufts University School of Medicine, Boston, USA; [b]Center for Evidence Synthesis in Health, School of Public Health, Brown University, Providence, RI; [c]Predictive Analytics and Comparative Effectiveness Center, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center/Tufts University School of Medicine, Boston, USA; [d]Center for the Evaluation of Value and Risk in Health, ICRHPS, Medical Center/Tufts University School of Medicine, Boston, USA; [e] VA Boston Healthcare System, Center for Healthcare Organization and Implementation Research (CHOIR); [f]National Pharmaceutical Council; [g]Department of Internal Medicine, University of California, Davis

Corresponding Author:
David Kent, MD, MS
Director, Predictive Analytics and Comparative Effectiveness (PACE) Center
Institute for Clinical Research and Health Policy Studies
Tufts Medical Center
800 Washington St, Box 63
Boston, MA 02111
Email: dkent1@tuftsmedicalcenter.org
Phone: 617-636-3234

Running title: Variation in person-level treatment effects: systematic review
Word count
Abstract: 224
Main text: 4,259 (main text, references)
Table: 5
Figures: 3
Key words: n-of-1 studies, systematic review, heterogeneity of treatment effect, personalized medicine

Abstract

Objective: Individual patients with the same condition may respond differently to similar treatments. Our aim is to summarize the reporting of person-level heterogeneity of treatment effects (HTE) in multi-person N-of-1 studies and to examine the evidence for person-level HTE through re-analysis.

Study Design: Systematic review and re-analysis of multi-person N-of-1 studies.

Data sources: Medline, Cochrane Controlled Trials, EMBASE, Web of Science, and review of references through August 2017 for N-of-1 studies published in English.

Study Selection: N-of-1 studies of pharmacological interventions with at least two subjects.

Data Synthesis: Citation screening and data extractions were performed in duplicate. We performed statistical reanalysis testing for person-level HTE on all studies presenting person-level data.

Results: We identified 62 multi-person N-of-1 studies with at least two subjects. Statistical tests examining HTE were described in only 13 (21%), of which only two (3%) tested person-level HTE. Only 25 studies (40%) provided person-level data sufficient to re-analyze person-level HTE. Reanalysis using a fixed effect linear model identified statistically significant person-level HTE in 8 of the 13 studies (62%) reporting person-level treatment effects and in 8 of the 14 studies (57%) reporting person-level outcomes.

Conclusions: Our analysis suggests person-level HTE is common and often substantial. Reviewed studies had incomplete information on person-level treatment effects and their variation. Improved assessment and reporting of person-level treatment effects in multi-person N-of-1 studies are needed.

Strengths and limitations of this study

- Multi-person N-of-1 studies are one of the best designs to estimate individual patient treatment effects and compare the variation in effects between individuals to variation within individuals across different periods
- This review highlights incomplete reporting of person-level treatment effects and their variation in multi-person N-of-1 studies.
- Re-analysis suggests person-level HTE is common and often substantial in multi-person N-of-1 studies, but varies from study to study.
- By distinguishing between condition-treatments with high versus low person-level HTE, multi-person N-of-1 studies have the potential to be important tools for personalized medicine.
- N-of-1 studies may be highly clinically informative for condition-treatments with a high degree of person-level HTE where the disease process is relatively stable over time, treatment effects are transient, and outcomes vary and are observable over time.

Introduction

Clinicians commonly observe that individual patients given the same treatment for the same condition frequently respond differently from one another. This observation, combined with our understanding of the complex mechanisms of diseases and therapies and the potential importance of myriad patient-specific factors (e.g., age, sex, illness severity, comorbidities, co-treatments, and molecular differences influencing pharmacokinetics and -dynamics), have led to a widely held assumption that the observed variation in treatment response seen between individuals is not merely random, but stable and potentially predictable. This assumption underpins the field of personalized medicine, which aims to determine the best treatment for an individual patient, as opposed to treating all patients with the same intervention found to be most effective for the "average" patient.

Nevertheless, statistical analyses aimed at discovering heterogeneity of treatment effects (HTE) among groups of individuals (for example subgroup analyses of parallel arm randomized trials) typically fail to find compelling and reliable evidence for the presence of such heterogeneity. For example, statistically significant differences in treatment effects between men and women are often reported, but a systematic review indicates that the frequency of these interactions across studies suggests the vast majority occur by chance.[1] Similarly, the field of pharmacogenetics, also built on the assumption of stable variation in treatment responses, has largely failed to live up to its promise to broadly improve the targeting of drugs—particularly outside the special case of oncology (where studies generally depend on the subclassification of tumor tissue not on variation in germline polymorphisms).[2;3] This failure to find reproducible HTE has supported the contrarian notion that true individual effects may be a "myth," an over-interpretation of random noise.[4]

To distinguish between these two possibilities, Kalow et al. have suggested that carefully designed series of N-of-1 studies could be performed for those chronic conditions amenable to this design (i.e., where the disease process is relatively stable over time, treatment effects are transient, and outcomes vary and are observable over time).[5] By estimating individual patient treatment effects and comparing the variation in effects between individuals to variation within individuals across different periods, it is possible to determine heterogeneity in individual treatment effects--even if one is unable to identify the variables that predict this variation (i.e., even in the absence of group-level HTE, such as men versus women, or old versus young).

A recent review summarized N-of-1 studies reported in the literature—including multi-person N-of-1 studies—but did not examine whether and how these studies provide information on person-level HTE. Therefore our objectives are: 1) to summarize the conduct and reporting of assessments of variation in person-level treatment effects from N-of-1 studies; and 2) to extract, reanalyze and report the results from the subset of studies that provided adequate data in their published reports to examine the extent of the evidence for person-level HTE (i.e., participant-level outcomes or effects).[6]

Methods

This review was conducted in accordance with the highest standards for conducing systematic reviews.[7;8] We defined N-of-1 studies as crossover trials in which each patient receives two or more treatments in a pre-defined, often randomized, sequence.

Data Sources and Searches

We used two separate searches because N-of-1 studies can be indexed differently: (1) a search in Medline, Cochrane Central and EMBASE using terms related to repeated crossover studies (for publications indexed from inception to August 17, 2017); and (2) a Medline, Cochrane Central, EMBASE, and Web of Science search using terms that are related to N-of-1 (for publications indexed from 2011 to August 17, 2017). For N-of-1 studies indexed before 2011, we used studies included in a prior published systematic review by Gabler et al.[6] Our searches combined terms and Medical Subject Headings for N-of-1, single-subject, single-patient, randomized trials, crossover, multi-period crossover, and rotated or repeated period crossover (see Appendix Tables 1-2 for detailed search terms). The searches were not restricted by disease, condition, organ system, or treatment.

Study Selection

We selected eligible multi-person N-of-1 studies to describe the frequency of reporting of individual outcomes and effects and of documented HTE in these studies. We required that a minimum of two individual subjects per study for evaluation of HTE. We excluded studies that included non-pharmacological interventions, reviews, abstracts and protocols. We include studies with placebo or "no treatment" interventions. Citations were double-screened by reviewers using an open-source, online software Abstrackr (http://abstrackr.cebm.brown.edu/). Full-text articles of potentially relevant studies were again double screened for eligibility.

Person-level outcomes were defined as outcomes for each person at each point in time when they were measured, reported in tables, text, or graphs. Person-level treatment effect was defined as contrasts of outcomes in individuals on one treatment versus the comparator. Person-level HTE was defined as quantified variation in the person-level treatment effects, whereas HTE

more broadly includes any type of subgroup analysis (e.g., males versus females; older versus younger) as outlined in **Figure 1**.

Data Extraction and Quality Assessment

One of four reviewers extracted data from each publication; a second reviewer verified all numerical information and basic descriptors of the study design and analysis. Operational definitions for extraction items were discussed in weekly project meetings and discrepancies between extractors were resolved by consensus with senior authors (DK, GR, EB). From each study, we extracted bibliographic information, details related to study design (number of patients enrolled, selection criteria, interventions evaluated, randomization methods, outcomes assessed, follow-up duration), information on patient characteristics, and person-level measurements of outcomes or estimates of person-level treatment effects (with corresponding measures of their uncertainty). When necessary, we extracted data by digitizing the graphs and the values were estimated using Engauge Digitizer version 2.14 (http://digitizer.sourceforge.net/). We assessed the methodological quality of each study based on predefined criteria, in accordance with the Agency for Healthcare Research and Quality (AHRQ) suggested methods and the Cochrane risk of bias for clinical trials.[9;10]

We generated graphs showing the trajectory of response for each patient in each study and compared them against the published information. We also generated scatterplots of measurements over time for studies that did not present their data in graphical format to help us identify aberrant data points (e.g., errors in data extraction). We verified potentially aberrant data points by re-examining the published data and made corrections, when needed.

Data Synthesis and Analyses

We examined the degree to which studies reported person-level data. This was described using the following items for each reported outcome: 1) qualitative descriptions of HTE (e.g., "there were 8 responders and 4 non-responders"); 2) details of person-level outcomes (i.e., outcomes with each treatment within each period); 3) details of person-level treatment effect (i.e., a point estimate of contrasts of outcomes in individuals on one treatment versus the comparator); 4) reporting of person-level statistical effect estimate, (e.g., standard deviation, exact P values, or confidence intervals for treatment effects within individuals); 5) description of statistical tests examining HTE (i.e., tests evaluating the contrast of treatment effects between individuals or groups in the study); and 6) claims of HTE. Note that qualitative descriptions of HTE for item 1 would include any description that implied that treatment effects varied, whereas item 6 required a more definite study conclusion (e.g., "our results demonstrate significant variation across individuals in response to treatment X"), whether or not these conclusions were based on robust statistical tests.

Statistical HTE analysis of extracted study results

We performed statistical analysis testing for person-level HTE on all studies presenting person-level data. We used a consistent analytic strategy across studies, to the extent permitted by the reporting in published papers. Our strategy was different for studies that reported person-level outcome measurements and those that reported estimates of person-level treatment effects with their sampling variances (or adequate information to approximately calculate these statistics).

For studies that only reported (or allowed the calculation of) *estimates of person-level treatment effects*, we obtained an average effect using a fixed effect inverse variance model and estimated the variance of the person-level treatment effects using DerSimonian and Laird method

of moments estimator.[11;12] In addition to a fixed effect model, we also obtained an average effect using a random effects model. Finally, we tested the hypothesis that all person-level treatment effects were equal using Cochran's chi-square test and quantified the proportion of observed variation due to "true" person-level effect heterogeneity with the I^2 statistic.[13]

For studies that reported *person-level outcomes*, we developed a linear model (for continuous outcomes) or generalized linear model (for binary or count outcomes) using the outcome of interest as the response, the intervention(s) as a covariate; indicator variables for different study participants were derived.[4] This model estimates a common treatment effect across participants. We also derived a similar model with treatment-by-participant interactions. This model allows each patient to have a different effect. The statistical significance of person-level HTE was assessed by a likelihood ratio test comparing the two models. In addition to a fixed effect model, we also fit a hierarchical linear or generalized linear mixed model with a random intercept and a random slope (for the treatment effect) to estimate the average treatment effect across all patients (assuming person-level HTE). We tested the hypothesis that all person-level treatment effects were equal and quantified the proportion of observed variation due to 'true' person-level effect heterogeneity with the I^2 statistic.[13]

Results

The searches for repeated crossover studies identified 11,891 citations and those for N-of-1 studies identified 3819 citations (indexed from 2011 onwards). Of these, we retrieved 407 full-text articles for review plus 100 N-of-1 trial articles (indexed before 2011) from an existing systematic review.[5] Upon full-text screening, 62 studies (58 multi-person N-of-1 studies and four repeated period crossover studies) met eligibility criteria (Appendix Table 3) and are reported

multi-person N-of-1 studies throughout the article. An outline of the search and study selection flow is provided in **Figure 2**.

Description of studies

Table 1 summarizes the 62 multi-person N-of-1 studies that were published between 1986 and 2017 reporting a total of 1974 patients. The most common clinical domains in the multi-person N-of-1 studies were neurology (16%), arthritis/rheumatology (10%) and psychiatry (9%). Most studies were described as "double-blind" but details about the methods for blinding were often unclear; similarly studies often provided unclear information about the generation of the randomization sequence and allocation concealment (Appendix Table 4). Among the studies, 93% compared a pair of treatment strategies, 5% compared three strategies, and 2% compared four strategies. Studies had between 3 and 16 treatment periods and obtained an average of 1 to 42 outcome measurements per period. Across reported outcomes, 89% of the assessed outcomes were patient-reported and 11% were investigator-assessed.

Reporting Person-level outcomes, effects and HTE

While most studies (92%) had some qualitative acknowledgement that the treatment effects appeared to vary across individuals, formal reporting at the participant level was variable **(Table 2)**. Person-level outcomes under each treatment were reported in 52% of multi-person N-of-1 studies. Person-level treatment effects with quantitative data (comparing outcomes on each treatment) for each individual who completed the trial was available in 32%; and details on the statistical evaluation of these effects (as standard deviations or exact P values or confidence intervals) were available in 13 (21%) multi-person N-of-1 studies. Only five (8%) studies

described statistical tests examining any HTE. However, only two studies (3%) reported person-level HTE, whereas the other two examined group-level HTE using conventional subgroup analysis based on observable characteristics.

Reanalysis of person-level data:

Of the 62 studies, there were 36 studies that provided person-level data, either as outcomes in each treatment period or as person-level treatment effects **(Table 3)**. Of these, only 25 studies provided person-level data sufficient to support re-analysis: 14 studies provided person-level outcomes; 13 studies provided person-level treatment effects (two studies provided both). The remaining 11 studies reported either medians or means without data on variance or did not provide sufficient information on completers, so they could not be re-analyzed for treatment effect or HTE.

Of 13 studies (with 27 unique comparisons) that reported analyzable person-level treatment effect data **(Table 3)**, 10 studies had a placebo comparator and three studies had an active comparator. The sample size ranged from 7 to 68; average crossover periods ranged from 6 to 16 days; and average outcome measures per period ranged from 1 to 21. The average treatment duration ranged from 14 to 336 days.

There were 14 studies (with 27unique comparisons) that reported analyzable person-level outcome data **(Table 3)**, including two studies also reporting person-level treatment effects. Of these, 11 compared the intervention with placebo and three studies compared two active interventions. The sample size ranged from 2 to 22; the average number of crossover periods ranged from 3 to 10; and the average number of outcome measures per period ranged from 1 to 42. The average treatment duration ranged from 9 to 210 days.

Re-analysis of studies reporting estimates of person-level treatment effects

Thirteen studies (including 27 comparisons, due to multiple outcomes in some studies) reported estimates of person-level treatment effects sufficient to analyze (Appendix Figures 1-16 displays graphs of the person level treatment effect data). Average fixed effect estimates for each analysis are shown in **Table 4**; random effects estimates were generally similar (Appendix 5). In 8 of the 13 studies (62%) and 15 of the 27 total unique comparisons (56%) we found evidence of statistically significant HTE for at least one outcome (Table 4). Generally, the magnitude in the variation of individual patient effects (as seen in the range) was very large compared to the average effects. Most studies (64%) showed person-level effects that differed qualitatively from one another. Most of the variation in the observed individual effects was attributable to "true" heterogeneity of person-level effects; 11 of 27 analyses had I^2 >80%.

Re-analysis of studies reporting person-level outcome measurements

Because some of the 14 studies providing analyzable outcome data had multiple outcomes (or multiple outcomes scales) there were a total of 27 comparisons with analyzable data. (Appendix Figures 17-42 displays graphs of the person level outcome results.) Average fixed effect estimates for each analysis are shown in Table 5; random effects estimates were generally similar (Appendix Table 6). In eight of the 14 studies (57%) (17 of the 27 unique comparisons [63%]), there was statistically significant person-level HTE for at least one outcome. Again, the variation in individual effects was often large compared to the average effect. However, given the lower number of participants per study and periods per participant and also different analytic approach, estimates of I^2 were much less precise in these studies.

Discussion

This review documents that multi-person N-of-1 studies rarely examine HTE. Only 8% of 62 multi-person N-of-1 studies described statistical tests examining HTE, but these generally involved comparisons of treatment effects among groups of patients (e.g., based on age or sex) rather than across individuals. Only two studies in the whole of the literature tested for person-level HTE.[14;15] Nevertheless, analyzable person-level results are sometimes reported in multi-person N-of-1 studies, as outcomes or as treatment effects. Our re-analyses of the totality of available data from these studies (n=25) suggested the presence of substantial variation in treatment effects across individuals in most studies. This was evident when considering statistical tests for the variation of treatment effects among patients and also by qualitative assessment of the magnitude of effect variation. This represents the first broad empirical examination with re-analysis of person-level HTE across multi-person N-of-1 studies, and it provides some general support for the *a priori* assumption of individual patient variation in treatment response that broadly motivates personalized medicine.

In contrast to parallel-group studies that establish efficacy in a group of patients with a common condition, N-of-1 studies establish the effects of an intervention in an individual.[16] In this respect, N-of-1 studies can be thought of as adjuncts to clinical care, where the goal is to select the right treatment for a particular patient, rather than as a research tool, where the goal is to create new generalizable knowledge.[17;18] Indeed, the results of traditional N-of-1 studies may be generalizable only to the future treatment response of the patient in the trial, not to other patients. Nevertheless, using Bayesian meta-analytic techniques, Zucker et al. showed how the average treatment effect at the population-level can also be estimated from combining multi-person N-of-1 studies testing similar interventions in similar patients with the same outcome

measures.[19] Similar Bayesian methods have also been suggested for analysis of group-level HTE.[20]

Herein, we demonstrate yet a new application of N-of-1 studies, to explore person-level HTE to describe the variation in individual treatment effects. This application has important research and clinical implications, even when the determinants of HTE remain unidentified. It is particularly of interest that there was apparent variation in the *degree* of person-level HTE found across conditions and treatments. Since the degree of variation across individuals sets the upper bound for the amount of HTE that might be explainable by observable characteristics, such as clinical or genomic variables, searching for subgroup effects in the absence of person-level HTE is a futile exercise.[4;21]

An interesting example of how person-level HTE can vary across different conditions comes from the study of Johannessen et al **(Figure 3)**.[14] These investigators conducted N-of-1 patient studies comparing cimetidine to placebo for patients presenting with dyspeptic symptoms and reported person-level effects by subgroups of disease categories. Among 46 trial completers, cimetidine had a significant effect for most patients (57%) and at the aggregate level. However, not only was there substantial person-level HTE, but person-level HTE varied across conditions, being much more pronounced in non-ulcer dyspepsia (I^2 = 75%) compared to peptic ulcer disease (I^2 = 35%) (Figure 3)— despite the very similar overall effects seen in these two conditions.

Finding variation in person-level response in multi-person N-of-1 studies identifies those conditions for which N-of-1 studies are likely to be clinically relevant. For condition-treatment combinations shown to have low person-level HTE, single subject studies are highly unlikely to be clinically informative, and the average results from trials (i.e., “one-size-fits-all” effects) are

more apt to be applicable to individuals.[22;23] On the other hand, N-of-1 studies may be highly clinically informative for condition-treatments with a high degree of person-level HTE. These conditions would also be potentially higher yield for examining predictors of HTE (genomic or otherwise).

Our findings also have implications for clinical practice and formulary design. For conditions marked by high person-level HTE, even when trials show that one treatment is better on average than others, having a variety of medication options would be useful to optimize outcomes across all patients, particularly for chronic conditions such as those studied here where empiric trials of alternative medications to find the best treatment for an individual might be feasible. For example, the study by March et al. shows that while patients with osteoarthritis on average had less pain and less stiffness with diclofenac, some patients had improved symptoms on paracetemol.[24] This person-level heterogeneity of treatment effect may not be detectable in conventional parallel arm trials employing conventional subgroup analysis.[21]

While more studies combining N-of-1 studies are needed to understand the extent of person-level HTE, future studies need to apply greater methodological rigor to improve the state-of-the-science on evaluation of individual treatment effects.[25] While the recently published CONSORT Extension for N-of-1 Trials may help improve reporting, a tabulation of all information (possibly electronically available) appears the most straightforward way to facilitate the clinical interpretation of these studies.[26] Such reporting allows the inspection of trajectories over time and may reveal patterns that are not captured by regression models. Complete reporting would also facilitate the development and evaluation of methods for the analysis of single subject experiments, particularly its use to better understand the extent and importance of person-level HTE.

The limitations of this review reflect, to a large extent, the limitations of the data in primary studies. Many conditions are not amenable to the N-of-1 design (e.g. because treatment effects are cumulative or because outcomes are observed only once). Further, even for conditions and treatment that are potentially amenable to this design, many important disease categories lacked published N-of-1 studies.. We relied on published studies only and our analytic cohort may be an underestimation of the true prevalence of these studies—particularly for N-of-1studies, which may frequently be conducted without the intention of future publication.

In addition, our conclusions regarding the ubiquity of HTE in the data we reanalyzed should be interpreted in the context of several important limitations. First, there were only a limited number of available studies that reported data sufficient to analyze, and therefore we present only a very partial picture of the full scope of inter-individual variation in effects across clinical conditions. Furthermore, among the studies that did have data, only fairly small numbers of patients were observed over a small number of treatment periods and we frequently had to rely on data summaries provided by the authors (e.g., person-level treatment effects and their sampling variance); these data limitations precluded the use of more complex models, for example models that account for period effects or other effects of time on the outcome.[3]

Our review has demonstrated that HTE remains almost totally unexplored in multi-person N-of-1 studies, which are uniquely capable of exploring variations in individual (person-level) treatment effects. Our re-analysis of the data from these studies represents the first systematic attempt to obtain empirical support for the *a priori* argument that treatment effects vary across individual patients, an assumption which underpins all efforts to personalize treatment selection. In this sample, person-level HTE appears to be fairly common and large enough to be clinically meaningful; the degree of person-level HTE appears to vary across conditions and outcomes.

Thus, multi-person N-of-1 studies are an under-utilized tool to identify where person-level HTE may be substantial, and where efforts to find molecular or clinical predictors of response heterogeneity should be focused. In such conditions, parallel arm studies might yield results that are over-generalized for patient level decision making.

Funding Source

This work was supported by the National Pharmaceutical Council.

Acknowledgement

We would like to acknowledge Issa Dahabreh, MD, MS, Assistant Professor of Health Services, Policy and Practice, Assistant Professor of Epidemiology, Brown University, for statistical advice.

We would like to acknowledge Tatum Williamson, MS, Research Assistant, Predictive Analytics and Comparative Effectiveness Center, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, for assistance with updating literature.

Contributorship statement

DK, GR made substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work. DK, GR are responsible for drafting the work or revising it critically for important intellectual content. DK, GR, EB, LL, JS, JC, JL, RD, RK have given final approval of the version to be published. DK, GR have made an agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Competing interests

None declared.

Data sharing statement

No additional data are available..

References

(1) Wallach JD, Sullivan PG, Trepanowski JF, Steyerberg EW, Ioannidis JP. Sex based subgroup differences in randomized controlled trials: empirical evidence from Cochrane meta-analyses. *BMJ* 2016; 355:i5826.

(2) Hertz DL, McLeod HL. Use of pharmacogenetics for predicting cancer prognosis and treatment exposure, response and toxicity. *J Hum Genet* 2013; 58(6):346-352.

(3) Kitsios GD, Kent DM. Personalised medicine: not just in our genes. *BMJ* 2012; 344:e2161.

(4) Senn S. Individual response to treatment: is it a valid assumption? *BMJ* 2004; 329(7472):966-968.

(5) Kalow W, Tang BK, Endrenyi L. Hypothesis: comparisons of inter- and intra-individual variations can substitute for twin studies in drug research. *Pharmacogenetics* 1998; 8(4):283-289.

(6) Gabler NB, Duan N, Vohra S, Kravitz RL. N-of-1 trials in the medical literature: a systematic review. *Med Care* 2011; 49(8):761-768.

(7) Finding What Works in Health Care: Standards for Systematic Reviews.Institute of Medicine (US) Committee on Standards for Systematic Reviews of Comparative Effectiveness Research. 2011.

(8) Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009; 151(4):264-9, W64.

(9) Methods Guide for Effectiveness and Comparative Effectiveness Review. AHRQ Publication No 10(11)-EHC063-EF Chapters 2011 Available from: URL:http://effectivehealthcare.ahrq.gov/

(10) Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011; 343:d5928.

(11) DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7(3):177-188.

(12) Schmidt FL, Oh IS, Hayes TL. Fixed- versus random-effects models in meta-analysis: model properties and an empirical comparison of differences in results. *Br J Math Stat Psychol* 2009; 62(Pt 1):97-128.

(13) Cochran WG. The combination of estimates from different experiments. Biometrics 10, 101-129. 1954.

(14) Johannessen T, Petersen H, Kristensen P, Fosstvedt D, Kleveland PM, Dybdahl J et al. Cimetidine on-demand in dyspepsia. Experience with randomized controlled single-subject trials. *Scand J Gastroenterol* 1992; 27(3):189-195.

(15) Pereira JA, Holbrook AM, Dolovich L, Goldsmith C, Thabane L, Douketis JD et al. Are brand-name and generic warfarin interchangeable? Multiple n-of-1 randomized, crossover trials. *Annals of Pharmacotherapy* 2005; 39(7-8):1188-1193.

(16) Guyatt GH, Heyting A, Jaeschke R, Keller J, Adachi JD, Roberts RS. N of 1 randomized trials for investigating new drugs. *Control Clin Trials* 1990; 11(2):88-100.

(17) Guyatt G, Sackett D, Taylor DW, Chong J, Roberts R, Pugsley S. Determining optimal therapy--randomized trials in individual patients. *N Engl J Med* 1986; 314(14):889-892.

(18) Guyatt G, Sackett D, Adachi J, Roberts R, Chong J, Rosenbloom D et al. A clinician's guide for conducting randomized trials in individual patients. *CMAJ* 1988; 139(6):497-503.

(19) Zucker DR, Ruthazer R, Schmid CH. Individual (N-of-1) trials can be combined to give population comparative treatment effect estimates: methodologic considerations. *J Clin Epidemiol* 2010; 63(12):1312-1323.

(20) Henderson NC, Louis TA, Wang C, Varadhan R. Bayesian analysis of heterogeneous treatment effects for patient-centered outcomes research. *Health Serv Outcomes Res Methodol* 2016; 16(4):213-233.

(21) Dahabreh IJ, Hayward R, Kent DM. Using group data to treat individuals: understanding heterogeneous treatment effects in the age of precision medicine and patient-centred evidence. *Int J Epidemiol* 2016;45(6):2184-2193.

(22) Simon G. Choosing a first-line antidepressant: equal on average does not mean equal for everyone. *JAMA* 2001; 286(23):3003-3004.

(23) Simon GE, Psaty BM, Hrachovec JB, Mora M. Principles for evidence-based drug formulary policy. *J Gen Intern Med* 2005; 20(10):964-968.

(24) March L, Irwig L, Schwarz J, Simpson J, Chock C, Brooks P. n of 1 trials comparing a non-steroidal anti-inflammatory drug with paracetamol in osteoarthritis. *BMJ* 1045; 309(6961):1041-1045.

(25) Kravitz RL, Duan N. eds, and the DEcIDE Methods Center N-of-1 Guidance Panel. Design and Implementation of N-of-1 Trials: A User's Guide. AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality. 2014.

(26) Vohra S, Shamseer L, Sampson M, Bukutu C, Schmid CH, Tate R et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. *BMJ* 2015; 350:h1738.

Table 1. Evidence Map of Multi-person N-of-1and Repeated Period Crossover Studies

Description	Multi-person N-of-1 Studies (n=62)
Publication Years	1979-2017
Subjects	**Total N (median, IQR)**
Enrolled	2153 (16, 9-42)
Completed	1705 (12, 7-32)
Intervention & Comparisons	
Head-to-head active drugs	10
Placebo	47
Active drug and placebo	1
Population	
Pediatric	12
Adults	50
Major Systems Studied	
Arthritis/Rheumatology	10
Cardiovascular	3
Gastrointestinal	7
Hypertension	1
Psychiatry	9
Neurology	16
Respiratory	9
Miscellaneous*	7
Top 5 Disease Conditions	
ADHD	6
Angina	3
Chronic Pain	5
GERD	5
Obstructive Airway	6
Osteoarthritis	6

**Sleep disorders, Allergy, Cancer, Muscular, Vascular (for multi-person N-of-1); Pain, Urology, GYN, , Heme/Onc, Allergy, Dermatology, Drug abuse, Endocrine, Lipids, Nephrology, Ophthalmology, Respiratory (for Repeated Cross-over Studies). ADHD, Attention-deficit hyperactivity disorder; GERD, Gastro-esophageal regurgitation disorder; IQR, Interquartile range; n, number of participants*

Table 2. Survey of HTE Assessment in Multi-person N-of-1 Studies

HTE Reporting	Multi-person N-of-1 Studies (n=62)
Qualitative description	92%
Person-level outcomes	52%
Person-level treatment effects	32%
Statistical analysis of person-level effects (e.g. p-values)	21%
Any statistical test for HTE	8%*
Claims of heterogeneity	15%

* Only 2 studies reported person-level HTE, the remaining 3 studies reported group level effect.

Table 3. Characteristics of studies reporting person-level data

Author, Year	Disease	Number enrolled (analyzed)	Intervention	Comparator	Cross-over periods	Total intervention duration	Outcome measures per period
Studies with re-analyzable person-level outcomes							
Camfield, 1996	Mental retardation with fragmented sleep	6 (6)	Melatonin	Placebo	7	10 wk	14
Hinderer, 1990	Traumatic spinal cord injury	5 (5)	Baclofen	Placebo	3	9 wk	2
Langer, 1993	Gastroesophageal reflux	2 (2)	Cisapride	Placebo	3	6 wk	5
Lashner, 1990	Ulcerative colitis	7 (6)	Nicotine	Placebo	4	8 wk	1
Maier, 1994	Chronic depression	10 (9)	Sulpiride	Placebo	4	28 wk	42
Mandelcorn, 2004	Brain injury	4 (4)	Ondansetron	Placebo	4	5 wk	1
McQuay, 1994	Neuropathic pain	19 (19)	Dextromethorphan	Placebo	5	20 d	1
Miyazaki, 1995	Unstable angina	22 (22)	Isosorbide dinitrate	Isosorbide dinitrate: intermittent injection	3	9 d	6
Nathan, 2006	Pediatric brain tumor	12 (7)	Ondansetron & metopimazine	Ondansetron & placebo	Unclear	189 d	unclear
Parodi, 1979	Unstable angina	12 (12)	Verapamil	Placebo	4	10 d	unclear
Parodi, 1986	Unstable angina	10 (10)	Verapamil	Propranolol, placebo	8	18 d	unclear
Tison, 2012	Levodopa-induced dyskinesia in Parkinson's disease patients	10 (10)	Simvastatin	Placebo	6	96 d	1
Studies with re-analyzable person-level treatment effects							
Emmanuel, 2012	Chronic intestinal pseudo-obstruction	7 (4)	Prucalopride	Placebo	16	48 wk	21
Haas, 2004	Chronic tension-type and migraine headache	39 (16)	Dextroamphetamine	Equi-stimulatory caffeine	8	20 d	20
Jaeschke, 1991	Fibromyalgia	22 (23)	Amitriptyline	Placebo	6	12 wk	2
Johannessen, 1992	Dyspepsia	68 (46)	Cimetidine	Placebo	12	184 d	15
Lipka, 2017	Autoimmune myasthenia	4 (4)	Ephedrine	Placebo	4	6 wk	1

	gravis						
Mahon, 1996	Irreversible chronic airflow limitation	16 (14)	Theophylline	Placebo	8	73 d	1
March, 1994	Osteoarthritis	25 (15)	Diclofenac	Paracetamol	6	12 wk	14
Patel, 1991	Nonreversible chronic airflow limitation	26 (18)	Ipratropium bromide / theophylline / salbutamol/ beclomethasone	Placebo	6	6 wk	Unclear
Wallace, 1994	Attention deficit hyperactivity disorder	11 (7)	Methylphenidate	Placebo	14	14 d	1
Woodfield, 2005	Skeletal muscle cramps	13	Quinine	Placebo	6	14 wk	2
Zucker, 2006	Fibromyalgia	58	Amitriptyline and Placebo	Amitriptyline and fluoxetine combination	6	36 wk	1
Study with both person-level data							
Pereira, 1995	Atrial fibrillation / deep venous thrombosis	7	Generic warfarin	Coumadin	10	30 wk	2
Joy, 2014	Statin-related myalgia	8 (7)	Statin	Placebo	6	33 wk	3
Study with insufficiently reported person-level data							
Person-level outcome data							
Denburg, 1994	Systemic lupus erythematosus	10	Prednisone	Placebo	6	30 wk	1
Mitchel, 2015	Fatigue in advanced cancer	43 (33)	Methylphenidate	Placebo	6	18 d	6
Nikles, 2000	Osteoarthritis	14	Ibuprofen	Paracetamol; Placebo	6	12 wk	14
Nikles, 2015	Dry mouth in advanced cancer	17 (4)	Pilocarpine	Placebo	6	18 d	6
Nikles, 2017	Acquired brain injury	53 (38)	Nervous system stimulants	Placebo	6	18 d	6
Reitberg, 2002	Allergic rhinitis	36	Loratadine and chlorpheniramine maleate	loratadine with placebo	8	32 d	4
Sheather-Reid, 1998	Chronic pain	8	Ibuprofen / Codeine	Placebo	6	12 wk	14
Person-level treatment effects							

Huber, 2007	Juvenile idiopathic arthritis	6	Amitriptyline	Placebo	6	17 wk	12
Privitera, 1994	Partial seizure	16	Dezinamide	Placebo	6	35 wk	6
Wegman, 2003	Osteoarthritis	13	Paracetamol	NSAIDs	10	20 wk	14
Wegman, 2005	Regular Temazepam users	15	Temazepam	Placebo	10	10 wk	7

Table 4. Analysis results of studies reporting person-level treatment effects

			Main Effect	*Person-Level Heterogeneity of Treatment Effect*		
Author Year	**Outcome**	**Range of the scales (severity)**	**Treatment effect (CI)**	**P for HTE***	**Treatment Effect Range**	**I-square % (CI)**
Emmanuel 2012	Bloating	0-4 (0=absent to 4=worst)	-0.344 (-0.619 to -0.069)	<0.001	-1.1 to -0.1	94 (88 to 97)
	Pain	0-4 (0=absent to 4=worst)	-0.440 (-0.771 to -0.110)	<0.001	-0.2 to -1.4	96 (92 to 98)
Haas 2004	Chronic tension-type headache grade	0-3 (0=none to 3=severe)	0.772 (0.454 to 1.090)	<0.001	0.04 to 1.9	84 (76 to 90)
	Chronic migraine headache grade	0-3 (0=none to 3=severe)	0.542 (0.354 to 0.731)	0.067	0.2 to 0.83	37 (0 to 65)
Jaeschke 1991	7-point symptom scale	1-7 (higher scores represent better function)	0.427 (0.210 to 0.645)	<0.001	-1.02 to 3.18	85 (79 to 89)
	Tender point changes count	Number of tender points	1.320 (0.404 to 2.236)	<0.001	-4.33 to 9.0	72 (57 to 82)
Johannessen 1992	6-point symptom scale	0-6 (0=NR to 6=NR)	0.698 (0.466 to 0.931)	<0.001	-1.67 to 3.17	66 (53 to 75)
Joy 2014	VAS myalgia Score	0-100mm (0=none to 100=worst)	0.119 (-2.283 to 2.521)	0.996	-8,10 to 9.45	0 (0 to 68)
	Symptom-specific VAS	0-100mm (0=none to 100=worst)	1.937 (0.179 to 3.696)	0.797	-8.0 to 18.05	0 (0 to 68)
	Pain severity score	0-10 (0=none to 10=worst)	0.086 (-0.215 to 0.387)	0.986	0.0 to 1.0	0 (0 to 68)
	Pain interference score	0-10 (0=none to 10=worst)	-0.016 (-0.095 to 0.064)	0.917	-0.02 to 0.75	0 (0 to 68)
Lipka 2017	Quantitative myasthenia gravis score	0-3 (0=none to 3=severe)	1.006 (0.215 to 1.797)	0.803	0.67 to 1.67	0 (0 to 85)
	Myasthenia gravis composite	0-50	2.891 (0.348 to 5.433)	0.177	-1.05 to 5.12	39 (0 to 80)
	MG-ADL	0-24	1.099 (-0.277 to 2.474)	0.047	0.03 to 3.0	62 (0 to 87)
	VAS score	0-10 (0=none to 100=worst)	1.275 (-0.115 to 2.665)	0.190	-0.01 to 3.02	37 (0 to 78)
Mahon 1996	Dyspnea in likert Scale	1-7 (1=extremely short of breath to 7=no shortness)	0.125 (-0.181 to 0.430)	<0.001	-0.57 to 0.89	78 (58 to 88)
March 1994	Mean pain score on VAS	5 point Likert scale (0-100mm)	-7.093 (-11.939 to -2.248)	<0.001	-33.8 to 4.1	98 (97 to 98)
	Mean stiffness score on VAS	5 point Likert scale (0-100mm)	-5.992 (-11.280 to -0.704)	<0.001	-36 to 10.7	97 (96 to 98)
Patel 1991**	4-item symptom questionnaire (All compared to placebo)	1-7 (1=extremely short of breath to 7=no shortness of breath)	0.240 (0.131 to 0.350)	<0.001	-0.34 to 3.1	91 (87 to 94)
	4-item symptom questionnaire (use of ipratropium bromide)		0.675 (0.264 to 1.085)	<0.001	-0.22 to 3.1	87 (78 to 92)
	4-item symptom questionnaire (use of salbutamol)		0.865 (0.042 to 1.687)	<0.001	0.46 to 1.3	94 (NA)
	4-item symptom questionnaire (use of theophylline)		0.025 (-0.434 to 0.484)	0.172	-0.34 to 0.18	30 (0 to 93)
Pereira	INR (diff)	Target INR range of 2.0–3.0	0.027 (-0.155 to 0.209)	0.477	-0.28 to 0.37	0 (0 to 75)

1995						
Wallace 1994	Conners 15-item rating scale scores	0-3 (NR)	0.759 (0.341 to 1.178)	0.747	0.42 to 1.22	0 (0 to 79)
Woodfield 2005	Changes in number of cramps	Number – mean difference	-18.823 (-28.527 to -9.120)	<0.001	-77 to -2	92 (87 to 95)
	Total days with cramps	days	-6.181 (-9.798 to -2.563)	<0.001	-13 to -1	94 (90 to 96)
Zucker 2006	FIQ	0-100 (0=best to 100=worst)	-5.019 (-8.784 to -1.254)	0.999	-32.0 to 0.98	0 (0 to 37)

* The significance of person-level HTE was assessed by Cochran's chi-square-based test
** One subject had beclomethasone

Table 5. Studies reporting person-level outcomes

Author Year	Outcome	Definition / Range of the Scales (severity)	*Main Effect*	*Person-level Heterogeneity of Treatment Effect*		
			Fixed Treatment Effect	P for Person Treatment Interaction*	Treatment Effect Range Lower Range (CI) Upper Range (CI)	I-square % (CI)
Camfield 1996	Nights without awakening	Between 10:00 PM and 7:00 AM per day	0.865 (0.215 to 1.516)	0.456	0.12 to 2.0	0 (0 to 79)
Hinderer 1990	Anxiety	Beck Inventory-A anxiety scale 0-3 (0 = never, 3 = almost all the time)	0.000 (0.000 to 0.000)	<0.001	-6.38 to 0.000	91 (81 to 95)
Joy 2014	Myalgia score	Visual Analogue Score for myalgia (0=none to 100=worst)	3.3812 (-2.668 to 9.430)	0.565	-11.66 to 60.79	0 (0 to 68)
Langer 1993	Vomiting	Number of episodes	-1.204 (-2.494 to 0.086)	0.136	-1.34 to 0.17	87 (NA)*
Lashner 1990	Symptom score: abdominal pain	Symptom scores 0-100 (0 = best, 100 = worst)	-3.615 (-16.982 to 9.751)	0.007	-35.0 to 15.0	37 (0 to 73)
	Symptom score: bowel movements/day		-0.538 (-1.215 to 0.138)	0.001	-3.0 to 1.0	56.6 (0 to 81)
	Symptom score: consistency of bowel movements		7.000 (-7.551 to 21.551)	0.013	-25.5 to 33.0	28 (0 to 69)
	Symptom score: hematochezia		2.308 (-17.210 to 21.826)	0.003	-38.0 to 47.5	47 (0 to 78)
	Symptom score: general sense of well-being		-6.538 (-25.352 to 12.275)	0.008	-43.0 to 35.0	35 (0 to 73)
Maier 1994	SCL-90 subscales: Depressed mood	Self-rating inventory to measure the effects of drug	-3.536 (-6.718 to -0.354)	<0.001	-17.8 to 2.74	58 (12 to 80)
	SCL-90 subscales: Anxiety		-3.753 (-6.582 to -0.924)	<0.001	-17.4 to 2.5	66 (30 to 83)
	SCL-90 subscales: Somatization		-1.419 (-4.316 to 1.478)	0.869	-6.0 to 2.7	0 (0 to 65)
Mandelcorn 2004	Self-Assessment score	0–5 (0 = worst, 5 = best)	-2.052 (-8.865 to 4.761)	0.05	-7.7 to 4.9	0 (0 to 85)
	Lower extremity ataxia	Fugl-Meyer: 3-point (0 cannot be performed to 2 can be fully performed)	12.494 (-3.155 to 28.142)	0.025	-6.42 to 36.76	35 (0 to 77)
	Truncal ataxia	AMTI forceplate®: NR *Berg Balance Scale® 0–56, with a higher score indicating a better performance*	1.196 (-2.866 to 5.257)	0.690	-0.52 to 2.20	0 (0 to 85)
	Upper extremity ataxia	Purdue Pegboard Test®: pegs inserted into the board with each hand in 30 sec	-0.498 (-3.546 to 2.550)	0.382	-3.68 to 1.42	0 (0 to 85)

Author Year	Outcome	Definition / Range of the Scales (severity)	*Main Effect* Fixed Treatment Effect	*Person-level Heterogeneity of Treatment Effect* P for Person Treatment Interaction*	Treatment Effect Range Lower Range (CI) Upper Range (CI)	I-square % (CI)
		Minnesota Placing Test®: reach out, grasp, and place blocks in a specific order				
McQuay 1994	VAS Pain Intensity	0-100 (0 = no pain, 100 = worst possible pain)	-1.094 (-5.572 to 3.383)	0.004	-8.0 to 10.1	0 (0 to 49)
	VAS Relief Intensity	0-100 (0 = no relief, 100 = complete pain relief)	-3.913 (-11.729 to 3.903)	0.038	-28.4 to 5.15	0 (0 to 49)
Miyazaki 1995	Incidence of angina	Either ST-segment elevation or depression at rest	0.496 (-0.206 to 1.199)	0.125	-16.19 to 17.11	0 (0 to 60)
Nathan 2006	Emetic episodes per day	complete response (0 episodes/day), major response (1–2 episodes/day), or failure (>2 episodes/day)	-0.095 (-0.514 to 0.325)	0.001	-16.5 to 2.08	59 (6 to 82)
Parodi 1979	Ischemic attacks	ST elevation or depression (details NR)	-1.544 (-1.838 to -1.251)	0.007	-16.21 to -0.34	48 (0 to 73)
Parodi 1986	Asymptomatic ST elevation (After verapamil)	0.1 mV of ST-segment elevation measured 20 ms after the J point	-1.637 (-1.994 to -1.279)	0.110	-2.37 to -1.30	6 (0 to 65)
	Asymptomatic ST depression (After verapamil)	More than 0.2 mV of ST-segment depression measured 80 ms after the J point	-1.083 (-1.903 to -0.262)	0.401	-17.42 to -0.90	0 (0 to 62)
	Symptomatic ST elevation (After verapamil)		-1.580 (-1.906 to -1.254)	<0.001	-15.40 to -1.45	0 (0 to 62)
	Symptomatic ST Depression (After verapamil)		-0.990 (-1.411 to -0.569)	0.002	-2.53 to -0.52	6 (0 to 64)
	Asymptomatic ST elevation (After propranolol)		0.100 (-0.086 to 0.286)	0.006	-0.77 to 1.38	62 (25 to 81)
	Asymptomatic ST depression (After propranolol)		0.339 (-0.168 to 0.845)	0.964	-18.3 to 0.83	0 (0 to 62)
	Symptomatic ST elevation (After propranolol)		-0.002 (-0.177 to 0.173)	0.063	-14.9 to 0.68	46 (0 to 74)
	Symptomatic ST Depression (After propranolol)		-0.374 (-0.709 to -0.039)	0.023	-17.1 to -0.73	4 (0 to 64)
Pereira 1995	INR	Target INR range of 2.0–3.0	-0.126 (-0.312 to 0.060)	0.433	-0.42 to 0.16	0 (0 to 71)
Tison 2012	Troublesome dyskinesia	7 points scale (1 = extremely uncomfortable, 7 = not at all uncomfortable)	0.167 (-0.449 to 0.783)	0.593	-0.67 to 1.83	0 (0 to 62)

* The significance of person-level HTE was assessed by a likelihood ratio test comparing the two models – model with common treatment effect and model with treatment-by-participant interactions

Figure Legend

Figure 1: The Figure provides a schematic description of: person-level outcomes (outcomes for each patient during each treatment period); person-level effects (contrasts of the outcomes for each patient in one treatment condition *versus* another); and person-HTE (between patient contrasts of effects).

Figure 2. Study Flow Diagram represents the flow of eligible studies included in this review

Figure 3. Person-level variation across different disease conditions. This figure depicts the results of 46 different N-of-1 trials of cimetidine as reported by Johanessen et al [12]. The effect of cimetidine versus placebo was measured in each subject across 12 cross-over periods over the span of 184 days. While cimetidine had a similar average effect regardless of the index condition, there was far greater consistency of effect in patients with peptic ulcer disease and much more variation in effect among patients with non-ulcer dyspepsia.

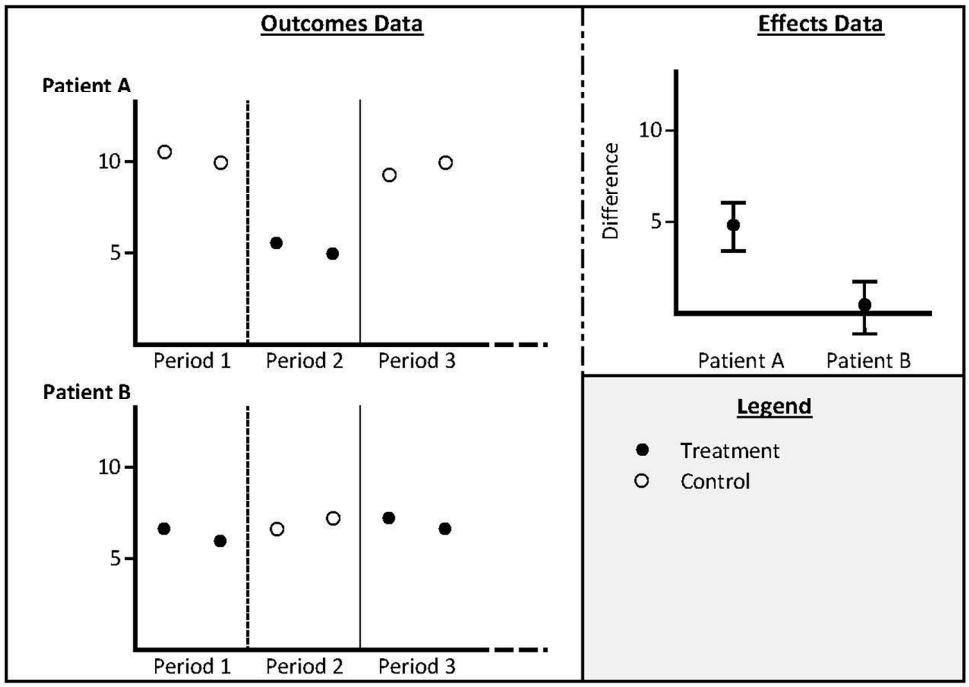


Figure 1: The Figure provides a schematic description of: person-level outcomes (outcomes for each patient during each treatment period); person-level effects (contrasts of the outcomes for each patient in one treatment condition versus another); and person-HTE (between patient contrasts of effects).

111x79mm (300 x 300 DPI)

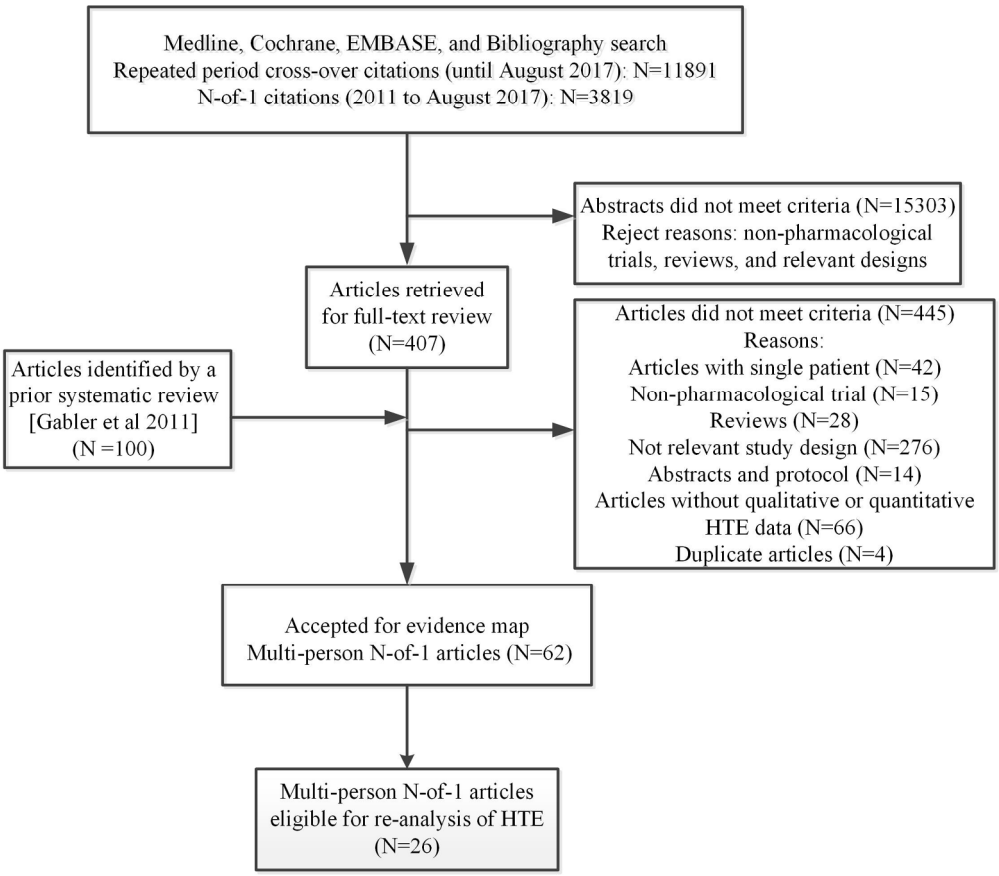


Figure 2. Study Flow Diagram represents the flow of eligible studies included in this review

189x164mm (300 x 300 DPI)

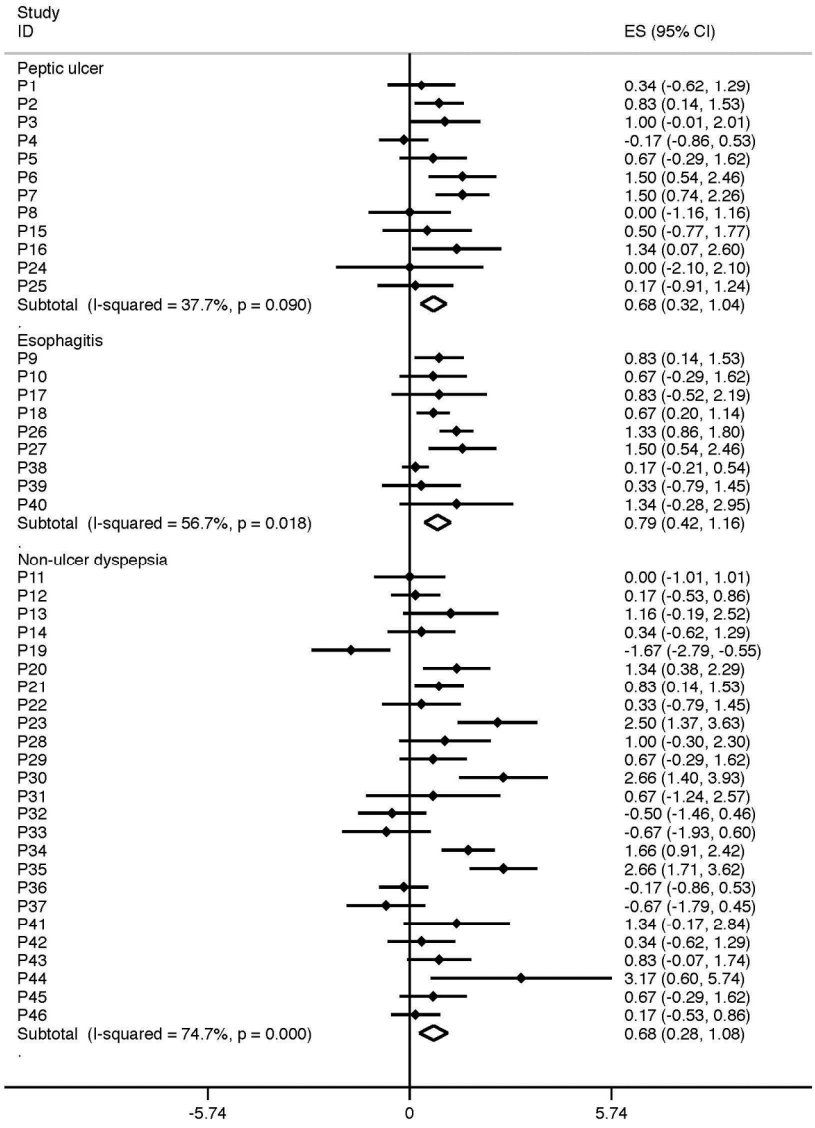


Figure 3. Person-level variation across different disease conditions. This figure depicts the results of 46 different N-of-1 trials of cimetidine as reported by Johanessen et al 12. The effect of cimetidine versus placebo was measured in each subject across 12 cross-over periods over the span of 184 days. While cimetidine had a similar average effect regardless of the index condition, there was far greater consistency of effect in patients with peptic ulcer disease and much more variation in effect among patients with non-ulcer dyspepsia.

150x203mm (300 x 300 DPI)

Appendix Materials

Appendix Table 1: N-of-1 Trial Searches

1.	randomized controlled trial.pt.
2.	controlled clinical trial.pt.
3.	randomized controlled trials/
4.	Double-blind Method/
5.	Single-Blind Method/
6.	clinical trial.pt.
7.	Clinical Trials.mp. or exp Clinical Trials/
8.	random$.tw.
9.	trial$.tw.
10.	Cross-Over Studies/
11.	or/1-10
12.	n-of-1.af.
13.	11 and 12
14.	(single-subject or single-patient or single case or single-case or within-patient).af.
15.	((single adj1 patient) or (single adj1 subject)).tw.
16.	14 or 15
17.	12 and 16
18.	multi-crossover.mp.
19.	12 and 18
20.	13 or 17 or 19
21.	limit 19 to yr="2010 - 2017"

Appendix Table 2: Repeated Period Crossover Trials

1.	(repeat$ or rotat$).af.
2.	((three or four or five or six) and period).tw.
3.	(multi- or multiple).tw.
4.	(three-period or four-period or five-period or six-period).tw.
5.	(three-way or four-way or five-way or six-way).tw.
6.	or/1-5
7.	Cross-Over Studies/ or (cross-over or crossover).af.
8.	6 and 7
9.	randomized controlled trial.pt.
10.	controlled clinical trial.pt.
11.	randomized controlled trials/
12.	Double-blind Method/
13.	Single-Blind Method/
14.	clinical trial.pt.
15.	Clinical Trials.mp. or exp Clinical Trials/
16.	random$.tw.
17.	trial$.tw.
18.	or/9-17
19.	8 and 18
20.	(dt or de or tu).fs.
21.	19 and 20
22.	7 and 20
23.	"Reproducibility of Results"/
24.	16 and 22
25.	limit 22 to english language
26.	9 or 10 or 11 or 14 or 15 or 16
27.	7 or 23
28.	20 and 26 and 27
29.	random.af.
30.	9 or 10 or 11 or 14 or 15 or 29
31.	ae.fs.
32.	20 or 31
33.	27 and 30 and 32
34.	limit 33 to (english language and humans)
35.	periods.af.

36.	6 or 35
37.	33 and 36
38.	Animals/ not human/
39.	37 not 38

Appendix Table 3: Reference List of Included Studies

1.	Nikles CJ, McKinlay L, Mitchell GK, Carmont SA, Senior HE, Waugh MC et al. Aggregated n-of-1 trials of central nervous system stimulants versus placebo for paediatric traumatic brain injury--a pilot study. Trials [Electronic Resource] 2014; 15:54.
2.	Tison F, Negre-Pages L, Meissner WG, Dupouy S, Li Q, Thiolat ML et al. Simvastatin decreases levodopa-induced dyskinesia in monkeys, but not in a randomized, placebo-controlled, multiple cross-over ("n-of-1") exploratory trial of simvastatin against levodopa-induced dyskinesia in Parkinson's disease patients. Parkinsonism & Related Disorders 2013; 19(4):416-421.
3.	Rascol O, Ferreira J, Negre-Pages L, Perez-Lloret S, Lacomblez L, Galitzky M et al. A proof-of-concept, randomized, placebo-controlled, multiple cross-overs (n-of-1) study of naftazone in Parkinson's disease. Fundamental & Clinical Pharmacology 2012; 26(4):557-564.
4.	Emmanuel AV, Kamm MA, Roy AJ, Kerstens R, Vandeplassche L. Randomised clinical trial: the efficacy of prucalopride in patients with chronic intestinal pseudo-obstruction--a double-blind, placebo-controlled, cross-over, multiple n = 1 study. Alimentary Pharmacology & Therapeutics 2012; 35(1):48-55.
5.	Yelland MJ, Poulos CJ, Pillans PI, Bashford GM, Nikles CJ, Sturtevant JM et al. N-of-1 randomized trials to assess the efficacy of gabapentin for chronic neuropathic pain. Pain Medicine 2009; 10(4):754-761.
6.	Nonoyama ML, Brooks D, Guyatt GH, Goldstein RS. Effect of oxygen on health quality of life in patients with chronic obstructive pulmonary disease with transient exertional hypoxemia. American Journal of Respiratory & Critical Care Medicine 2007; 176(4):343-349.
7.	Huber AM, Tomlinson GA, Koren G, Feldman BM. Amitriptyline to relieve pain in juvenile idiopathic arthritis: a pilot study using Bayesian metaanalysis of multiple N-of-1 clinical trials. Journal of Rheumatology 2007; 34(5):1125-1132.
8.	Yelland MJ, Nikles CJ, McNairn N, Del Mar CB, Schluter PJ, Brown RM. Celecoxib compared with sustained-release paracetamol for osteoarthritis: a series of n-of-1 trials. Rheumatology 2007; 46(1):135-140.
9.	Zucker DR, Ruthazer R, Schmid CH, Feuer JM, Fischer PA, Kieval RI et al. Lessons learned combining N-of-1 trials to assess fibromyalgia therapies. Journal of Rheumatology 2006; 33(10):2069-2077.
10.	Nikles CJ, Mitchell GK, Del Mar CB, Clavarino A, McNairn N. An n-of-1 trial service in clinical practice: testing the effectiveness of stimulants for attention-deficit/hyperactivity disorder. Pediatrics 2006; 117(6):2040-2046.
11.	Nathan PC, Tomlinson G, Dupuis LL, Greenberg ML, Ota S, Bartels U et al. A pilot study of ondansetron plus metopimazine vs. ondansetron monotherapy in children receiving highly emetogenic chemotherapy: a Bayesian randomized serial N-of-1 trials design. Supportive Care in Cancer 2006; 14(3):268-276.
12.	Pereira JA, Holbrook AM, Dolovich L, Goldsmith C, Thabane L, Douketis JD et al. Are brand-name and generic warfarin interchangeable? Multiple n-of-1 randomized, crossover trials. Annals of Pharmacotherapy 2005; 39(7-8):1188-1193.
13.	Woodfield R, Goodyear-Smith F, Arroll B. N-of-1 trials of quinine efficacy in skeletal muscle cramps of the leg. British Journal of General Practice 2005; 55(512):181-185.
14.	Wegman AC, van der Windt DA, Bongers M, Twisk JW, Stalman WA, de Vries TP. Efficacy of temazepam in frequent users: a series of N-of-1 trials. Family Practice 2005; 22(2):152-159.
15.	Nikles CJ, Yelland M, Glasziou PP, Del MC. Do individualized medication effectiveness tests (n-of-1 trials) change clinical decisions about which drugs to use for osteoarthritis and chronic pain?. [Review] [19 refs]. American Journal of Therapeutics 2005; 12(1):92-97.
16.	Smith BJ, Appleton SL, Veale AJ, McElroy HJ, Veljkovic D, Saccoia L. Eformoterol n-of-1 trials in chronic obstructive pulmonary disease poorly reversible to salbutamol. Chronic Respiratory Disease 2004; 1(2):63-69.

17.	Haas DC, Sheehe PR. Dextroamphetamine pilot crossover trials and n of 1 trials in patients with chronic tension-type and migraine headache. Headache 2004; 44(10):1029-1037.
18.	Mandelcorn J, Cullen NK, Bayley MT. A preliminary study of the efficacy of ondansetron in the treatment of ataxia, poor balance and incoordination from brain injury. Brain Injury 2004; 18(10):1025-1039.
19.	Pope JE, Prashker M, Anderson J. The efficacy and cost effectiveness of N of 1 studies with diclofenac compared to standard treatment with nonsteroidal antiinflammatory drugs in osteoarthritis. Journal of Rheumatology 2004; 31(1):140-149.
20.	Wegman AC, van der Windt DA, de HM, Deville WL, Fo CT, de Vries TP. Switching from NSAIDs to paracetamol: a series of n of 1 trials for individual patients with osteoarthritis. Annals of the Rheumatic Diseases 2003; 62(12):1156-1161.
21.	Reitberg DP, Del RE, Weiss SL, Rebell G, Zaias N. Single-patient drug trial methodology for allergic rhinitis. Annals of Pharmacotherapy 2002; 36(9):1366-1374.
22.	Linday LA, Tsiouris JA, Cohen IL, Shindledecker R, DeCresce R. Famotidine treatment of children with autistic spectrum disorders: pilot research using single subject research design. Journal of Neural Transmission 2001; 108(5):593-611.
23.	Duggan CM, Mitchell G, Nikles CJ, Glasziou PP, Del Mar CB, Clavarino A. Managing ADHD in general practice. N of 1 trials can help! Australian Family Physician 2000; 29(12):1205-1209.
24.	Nikles CJ, Glasziou PP, Del Mar CB, Duggan CM, Clavarino A, Yelland MJ. Preliminary experiences with a single-patient trials service in general practice. Medical Journal of Australia 2000; 173(2):100-103.
25.	Mahon JL, Laupacis A, Hodder RV, McKim DA, Paterson NA, Wood TE et al. Theophylline for irreversible chronic airflow limitation: a randomized study comparing n of 1 trials to standard practice. Chest 1999; 115(1):38-48.
26.	Bollert FG, Paton JY, Marshall TG, Calvert J, Greening AP, Innes JA. Recombinant DNase in cystic fibrosis: a protocol for targeted introduction through n-of-1 trials. Scottish Cystic Fibrosis Group. European Respiratory Journal 1999; 13(1):107-113.
27.	Kent MA, Camfield CS, Camfield PR. Double-blind methylphenidate trials: practical, useful, and highly endorsed by families. Archives of Pediatrics & Adolescent Medicine 1999; 153(12):1292-1296.
28.	Webb S, Tansey P, Brown H, Jackson A, Bilton D. Placebo-controlled n-of-1 trials in cystic fibrosis. European Respiratory Journal 1999; 14(4):993.
29.	Haines DR, Gaines SP. N of 1 randomised controlled trials of oral ketamine in patients with chronic pain. Pain 1999; 83(2):283-287.
30.	Sheather-Reid RB, Cohen M. Efficacy of analgesics in chronic pain: a series of N-of-1 studies. Journal of Pain & Symptom Management 1998; 15(4):244-252.
31.	Camfield P, Gordon K, Dooley J, Camfield C. Melatonin appears ineffective in children with intellectual deficits and fragmented sleep: six "N of 1" trials. Journal of Child Neurology 1996; 11(4):341-343.
32.	Mahon J, Laupacis A, Donner A, Wood T. Randomised study of n of 1 trials versus standard practice.[Erratum appears in BMJ 1996 Jun 1;312(7043):1392]. BMJ 1996; 312(7038):1069-1074.
33.	Maier W, Benkert O. Treatment of chronic depression with sulpiride: evidence of efficacy in placebo-controlled single case studies. Psychopharmacology 1994; 115(4):495-501.
34.	McQuay HJ, Carroll D, Jadad AR, Glynn CJ, Jack T, Moore RA et al. Dextromethorphan for the treatment of neuropathic pain: a double-blind randomised controlled crossover trial with integral n-of-1 design. Pain 1994; 59(1):127-133.
35.	March L, Irwig L, Schwarz J, Simpson J, Chock C, Brooks P. n of 1 trials comparing a non-steroidal anti-inflammatory drug with paracetamol in osteoarthritis. BMJ 1994; 309(6961):1041-1045.
36.	Denburg SD, Carbotte RM, Denburg JA. Corticosteroids and neuropsychological functioning in

	patients with systemic lupus erythematosus. Arthritis & Rheumatism 1994; 37(9):1311-1320.
37.	Privitera MD, Treiman DM, Pledger GW, Sahlroot JT, Handforth A, Linde MS et al. Dezinamide for partial seizures: results of an n-of-1 design trial. Neurology 1994; 44(8):1453-1458.
38.	Langer JC, Winthrop AL, Issenman RM. The single-subject randomized trial. A useful clinical tool for assessing therapeutic efficacy in pediatric practice. Clinical Pediatrics 1993; 32(11):654-657.
39.	Molloy DW, Guyatt GH, Standish T, Willan A, McIlroy W, D'Souza J et al. Effect of a new nootropic agent, CGS 5649B, on cognition, function, and behavior in dementia. Journal of General Internal Medicine 1993; 8(8):444-447.
40.	Johannessen T, Petersen H, Kristensen P, Fosstvedt D, Kleveland PM, Dybdahl J et al. Cimetidine on-demand in dyspepsia. Experience with randomized controlled single-subject trials. Scandinavian Journal of Gastroenterology 1992; 27(3):189-195.
41.	Johannessen T, Kristensen P, Petersen H, Fosstvedt D, Loge I, Kleveland PM et al. The symptomatic effect of 1-day treatment periods with cimetidine in dyspepsia. Combined results from randomized, controlled, single-subject trials. Scandinavian Journal of Gastroenterology 1991; 26(9):974-980.
42.	Patel A, Jaeschke R, Guyatt GH, Keller JL, Newhouse MT. Clinical usefulness of n-of-1 randomized controlled trials in patients with nonreversible chronic airflow limitation. American Review of Respiratory Disease 1991; 144(4):962-964.
43.	Larsen S, Farup P, Flaten O, Osnes M. The multi-crossover model for classifying patients as responders to a given treatment. Scandinavian Journal of Gastroenterology 1991; 26(7):763-770.
44.	Jaeschke R, Adachi J, Guyatt G, Keller J, Wong B. Clinical usefulness of amitriptyline in fibromyalgia: the results of 23 N-of-1 randomized controlled trials. Journal of Rheumatology 1991; 18(3):447-451.
45.	Hinderer SR. The supraspinal anxiolytic effect of baclofen for spasticity reduction. American Journal of Physical Medicine & Rehabilitation 1990; 69(5):254-258.
46.	Lashner BA, Hanauer SB, Silverstein MD. Testing nicotine gum for ulcerative colitis patients. Experience with single-patient trials. Digestive Diseases & Sciences 1990; 35(7):827-832.
47.	McBride MC. An individual double-blind crossover trial for assessing methylphenidate response in children with attention deficit disorder. Journal of Pediatrics 1988; 113(1:Pt 1):t-45.
48.	Menard J, Serrurier D, Bautier P, Plouin PF, Corvol P. Crossover design to test antihypertensive drugs with self-recorded blood pressure. Hypertension 1988; 11(2):153-159.
49.	Ullmann RK, Sleator EK. Responders, nonresponders, and placebo responders among children with attention deficit disorder. Importance of a blinded placebo evaluation. Clinical Pediatrics 1986; 25(12):594-599.
50.	Wolfe B, Del RE, Weiss SL, Mendelson A, Elbaga TA, Huser FJ et al. Validation of a single-patient drug trial methodology for personalized management of gastroesophageal reflux disease. J Manag Care Pharm 2002; 8(6):459-468.
51.	Brookes ST, Biddle L, Paterson C, Woolhead G, Dieppe P. "Me's me and you's you": Exploring patients' perspectives of single patient (n-of-1) trials in the UK. Trials 2007; 8:10.
52.	Wallace AE, Kofoed LL. Statistical Analysis of Single Case Studies in the Clinical Setting: The Example of Methylphenidate Trials in Children with Attention-Deficit Hyperactivity Disorder. Journal of Child and Adolescent Psychopharmacology 1994; 4(3):141-150.
53.	Miyazaki S, Nonogi H, Goto Y, Sumiyoshi T, Haze K, Hiramori K. Comparison of the therapeutic efficacy of continuous and intermittent injection of isosorbide dinitrate: a randomized study on unstable angina. Internal Medicine 1995; 34(9):856-862.
54,	Parodi O, Maseri A, Simonetti I. Management of unstable angina at rest by verapamil. A double-blind cross-over study in coronary care unit. British Heart Journal 1979; 41(2):167-174.
55.	Parodi O, Simonetti I, Michelassi C, Carpeggiani C, Biagini A, L'Abbate A et al. Comparison of verapamil and propranolol therapy for angina pectoris at rest: a randomized, multiple-crossover, controlled trial in the coronary care unit. American Journal of Cardiology 1986; 57(11):899-906.

56.	Joy TR, Zou GY, Mahon JL. N-of-1 (single-patient) trials for statin-related myalgia. Annals of internal medicine 2014 Oct 7;161(7):531-2
57.	Lipka AF, Vrinten C, van Zwet EW, Schimmel KJM, Cornel MC, Kuijpers MR, et al. Ephedrine treatment for autoimmune myasthenia gravis. Neuromuscular disorders 2017; 27:259-265.
58.	McGarry ME, Illek B, Ly NP, Zlock L, Olshansky S, Moreno C, et al. In vivo and in vitro ivacaftor response in cystic fibrosis patients with residual CFTR function: n-of-1 studies. Pediatric pulmonology 2017;52(4):472-9.
59.	Mitchell GK, Hardy JR, Nikles CJ, Carmont SA, Senior HE, Schluter PJ, et al. The Effect of Methylphenidate on Fatigue in Advanced Cancer: An Aggregated N-of-1 Trial. Journal of pain and symptom management 2015 Sep;50(3):289-96.
60.	Nikles J, Mitchell GK, Hardy J, Agar M, Senior H, Carmont SA, et al. Testing pilocarpine drops for dry mouth in advanced cancer using n-of-1 trials: A feasibility study. Palliative Medicine 2015 Dec;29(10):967-74.
61.	Nikles J, Mitchell GK, Hardy J, Senior H, Carmont SA, Schluter PJ, et al. Single-patient multiple crossover studies to determine the effectiveness of paracetamol in relieving pain suffered by patients with advanced cancer taking regular opioids: A pilot study. 2016;30(8):800-2.
62.	Nikles J, Mitchell G, McKinlay L, Waugh MC, Epps A, Carmont SA, et al. A series of n-of-1 trials of stimulants in brain injured children. 2017;40(1):11-21.

Appendix Table 4: Risk of bias assessment

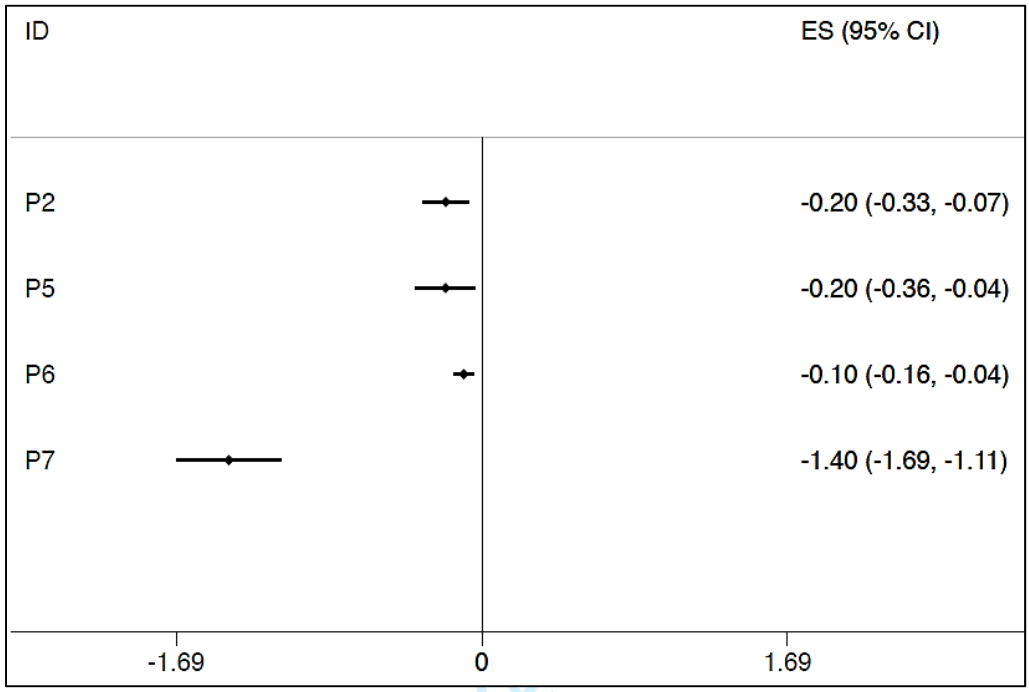
Author Yr	1. Randomization adequate?	2. Allocation concealed?	3. Patient blinded?	4. Outcome assessor blinded?	5. run-in period?	7. Wash-out?	8. Statistical methods appropriate?*	9. All randomized participants analyzed?	10. Incomplete outcome data
Nikles 2014	Low	Low	Low	Low	High	High	Low	High	Low
Tison 2013	Unclear	Low	Low	Low	High	Low	High	Low	Low
Rascol 2012	Unclear	Unclear	Low	Low	Low	Low	Low	High	Low
Emmanuel 2012	Unclear	Unclear	Low	Low	High	High	High	High	Low
Yelland 2009	Low	Low	Low	Low	High	High	Low	High	Low
Brookes 2007	Low	Low	Low	Low	High	High	unclear	High	Low
Nonoyama2007	Low	Low	Low	Low	High	High	unclear	High	Low
Huber 2007	Low	Low	Low	Low	High	Low	High	Low	Low
Yelland 2007	Low	Unclear	Low	Low	High	High	Low	High	Low
Zucker 2006	Low	Unclear	Low	Low	Low	High	Low	High	Low
Nikles 2006	Low	Unclear	Low	Low	High	Low	High	High	Low
Nathan 2006	Low	Low	Low	Low	High	High	High	High	Low
Pereira 1995	Unclear	Low	Low	Low	High	High	High	Low	Low
Woodfield 2005	Low	Low	Low	Low	Low	Low	Low	Low	Low
Wegman 2005	Low	Unclear	Low	Low	High	High	Low	High	Low
Nikles 2005	Low	Unclear	Low	Low	High	Low	High	High	Low
Smith 2004	Low	Low	Low	Low	Low	High	Low	High	Low
Haas 2004	Low	Low	Low	Low	High	High	Low	High	Low
Mandelcorn 2004	Low	Unclear	Low	Low	Low	High	High	Low	Low
Pope 2004	Unclear	High	Low	Low	High	High	Low	Low	Low
Wegman 2003	Low	Low	Low	Low	High	High	Low	High	Low

Wolfe 2002	Low	Low	Low	Low	High	Low	Low	High	Low
Reitberg 2002	Low	Low	Low	Low	Low	High	Low	Low	Low
Linday 2001	Unclear	Low	Low	Low	Low	High	High	High	Low
Duggan 2000	Unclear	Unclear	Low	Low	High	High	High	Low	Low
Nikles 2000	Unclear	Unclear	Low	Low	High	High	High	High	Low
Mahon 1999	Low	Unclear	Low	Low	High	Low	High	High	Low
Bollert 1999	Unclear	Unclear	Low	Low	High	High	High	High	Low
Kent 1999	Unclear	Unclear	Low	Low	High	High	High	Low	Low
Webb 1999	Unclear	Unclear	Low	Low	Low	Low	High	High	Low
Haines 1999	Unclear	Unclear	Low	Low	Low	Low	Low	High	Low
Sheather-Reid 1998	Unclear	Unclear	Low	Low	High	High	Low	High	Low
Camfield 1996	Unclear	Unclear	Low	Low	High	High	High	Low	Low
Mahon 1996	Low	Unclear	Low	Low	High	High	High	Low	Low
Miyazaki 1995	Unclear	High	High	High	High	High	High	High	Low
Maier 1994	Unclear	Unclear	Low	Low	Low	Low	Low	High	Low
McQuay 1994	Low	Low	Low	Low	High	High	High	High	Low
March 1994	Unclear	Unclear	Low	Low	High	High	Low	High	Low
Denburg 1994	Unclear	Unclear	Low	Low	Low	Low	High	High	Low
Privitera 1994	Unclear	Unclear	Low	Low	High	High	High	Low	Low
Wallace 1994	High	Unclear	Low	Low	High	High	High	Low	Low
Langer 1993	Low	Low	Low	Low	High	High	High	Low	Low
Molloy 1993	Unclear	Unclear	Low	Low	Low	Low	Low	High	Low
Johannessen 1992	Unclear	Unclear	Low	Low	High	Low	High	High	Low
Johannessen 1991	Unclear	Unclear	Low	Low	High	High	Low	Low	Low
Patel 1991	Unclear	Unclear	Low	Low	High	High	High	Low	Low
Larsen 1991	Unclear	Unclear	Low	Low	High	High	High	High	Low
Jaeschke 1991	Unclear	Unclear	Low	Low	low	High	High	High	low
Hinderer 1990	Unclear	Unclear	Low	Low	low	High	high	low	low
Lashner 1990	Unclear	Low	Low	Low	Unclear	High	high	low	low

McBride 1988	Low	Low	Low	Low	Unclear	High	high	low	High
Menard 1988	Low	Unclear	Low	Low	Low	low	low	low	High
Ullmann 1986	Low	Unclear	Low	Low	Unclear	High	low	low	High
Parodi 1986	Low	Unclear	Low	Low	low	Low	low	low	low
Parodi 1979	Unclear	Unclear	Low	Low	low	High	High	low	low
Joy 2014	Low	Unclear	Low	Low	High	Low	low	low	low
Lipka 2017	Low	Low	Low	Low	High	Low	High	low	low
Mitchell 2015	Low	Low	Low	Low	High	Low	High	low	low
Nikles 2015	Low	Low	Low	Low	High	Low	High	low	low
Nikles 2017	Low	Low	Low	Low	High	Low	low	High	low
Nikles 2016	Low	Unclear	Low	Low	High	High	High	low	High
McGarry 2017	Low	Low	Low	Low	Low	Low	High	High	High

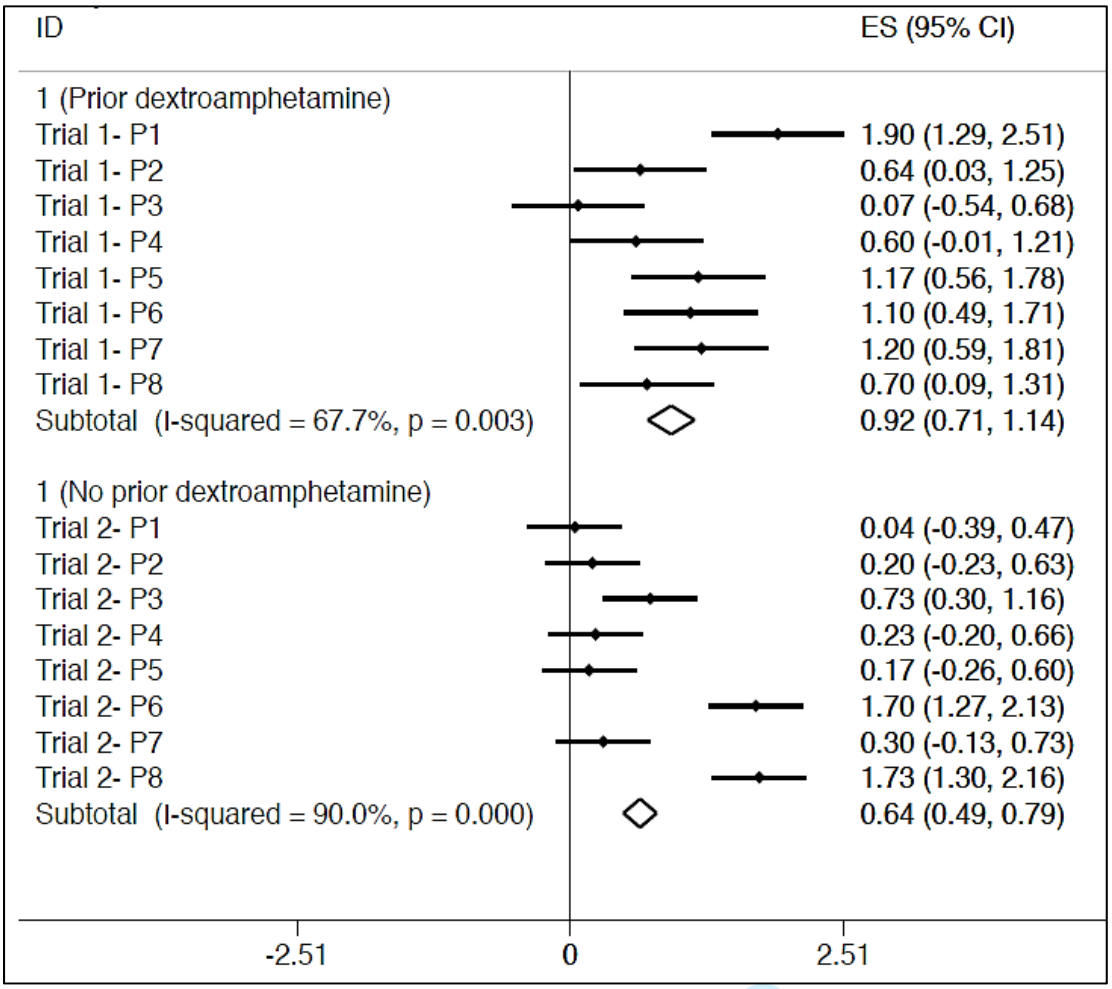
* Statistical methods used to account for carryover effect, period effects, and intra-subject correlation

Appendix Figure 1: Patients with chronic intestinal pseudo-obstruction treated with prucalopride or placebo for pain relief[1]



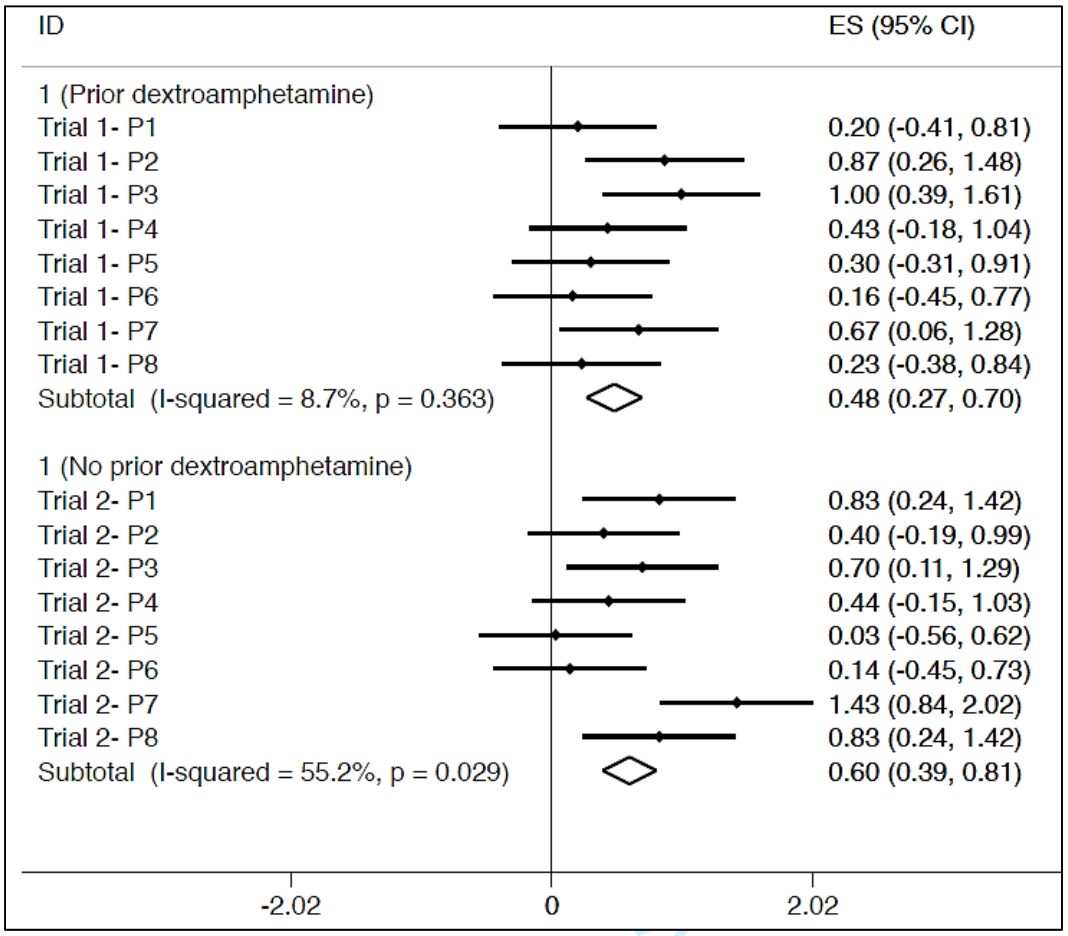
Appendix Figure 1 Legend:
Data from this figure was extracted from the study published by Emmanuel et al in 2011, which investigates the use of prucalopride or placebo for pain relief (among other outcomes) in patients with chronic intestinal pseudo-obstruction. The average treatment effect is -0.440 (-0.771 to -0.110).

Appendix Figure 2: Patients with chronic tension-type headaches treated with dextroamphetamine or control and effect on mean daily grade decrease in headache[2]



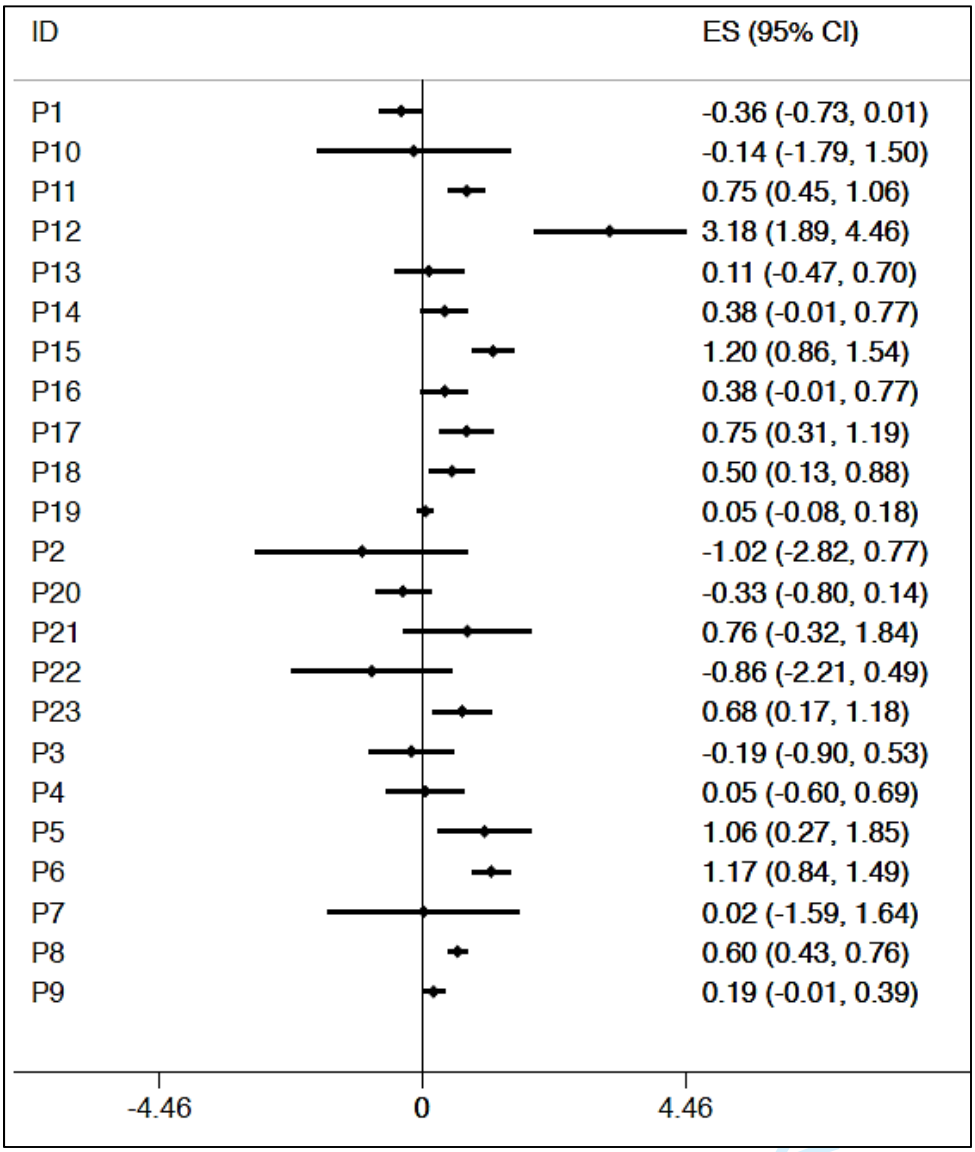
Appendix Figure 2 Legend: Data from this figure was extracted from the study published by Haas et al in 2004, which investigates the use of dextroamphetamine or control in patients with chronic-type for improvement on mean daily grade in headache.

Appendix Figure 3: Patients with migraine headaches treated with dextroamphetamine or control and effect on mean daily grade decrease in headache[2]



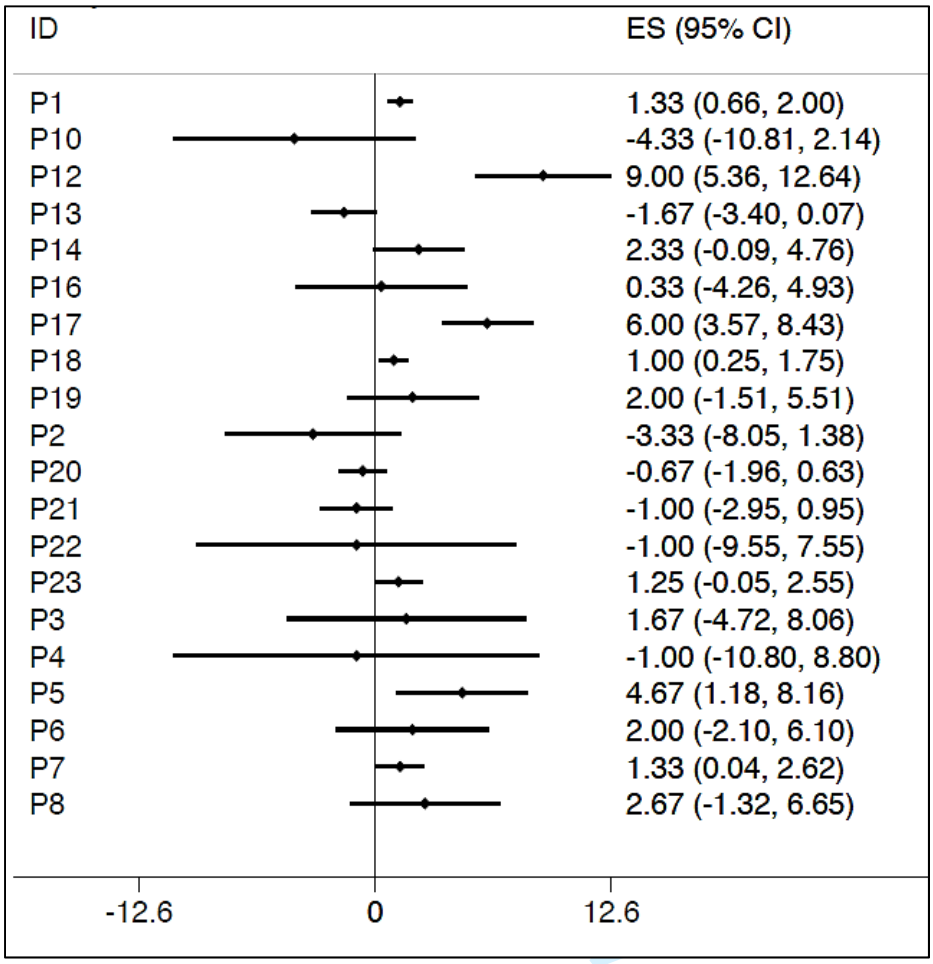
Appendix Figure 3 Legend: Data from this figure was extracted from the study published by Haas et al in 2004, which investigates the use of dextroamphetamine or control in patients with chronic-type and migraine headaches for improvement on mean daily grade in headache.

Appendix Figure 4: Patients with fibromyalgia treated with amitriptyline or placebo and its effect on a 7-point symptom scale[3]



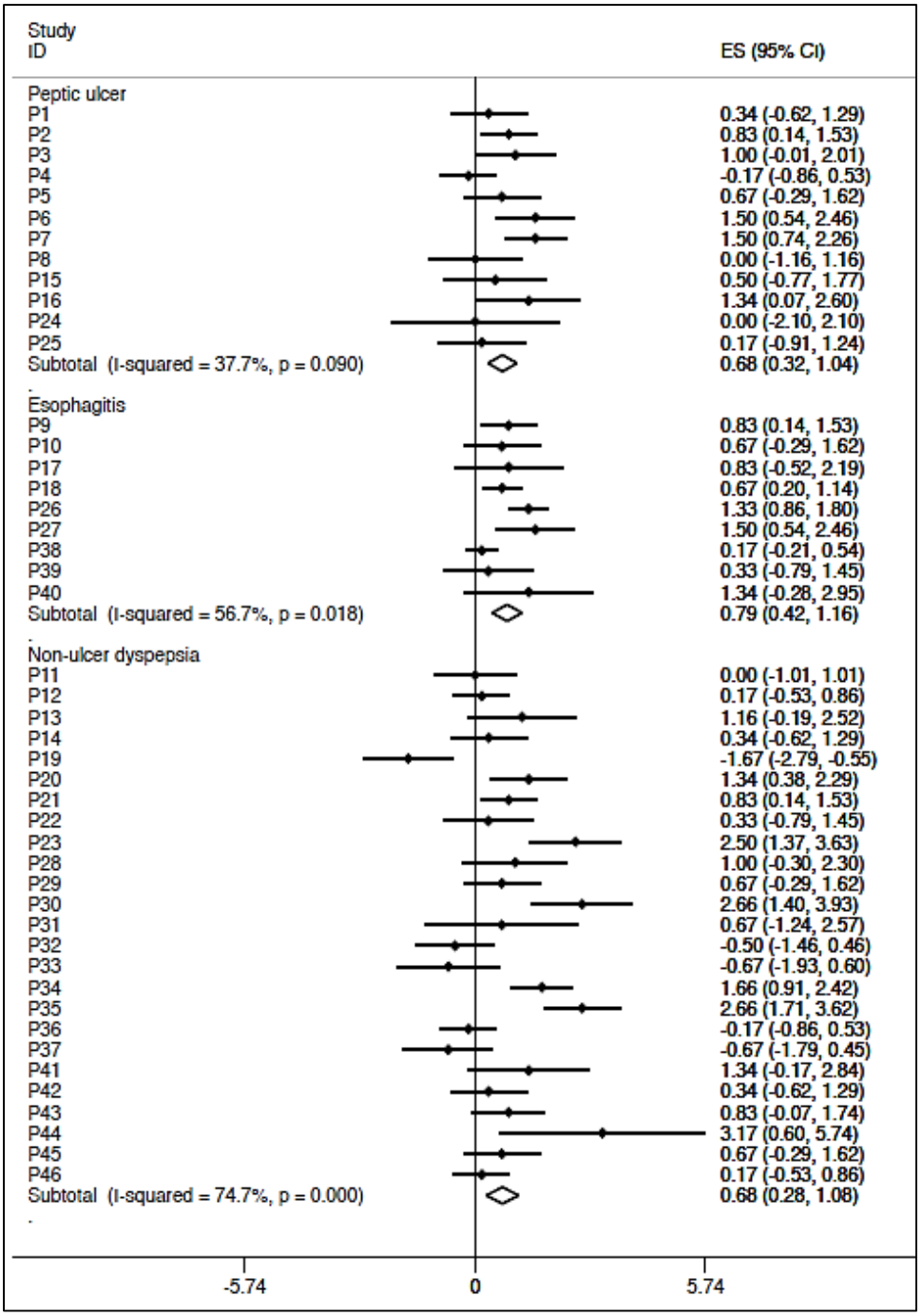
Appendix Figure 4 Legend: Data from this figure was extracted from the study published by Jaeschke et al in 1991, which investigates the effect of amitriptyline or placebo on a 7-point symptom scale in patients with fibromyalgia. The average treatment effect is 0.427 (0.210 to 0.645).

Appendix Figure 5: Patients with fibromyalgia treated with amitriptyline or placebo and its effect on tender point changes count[3]



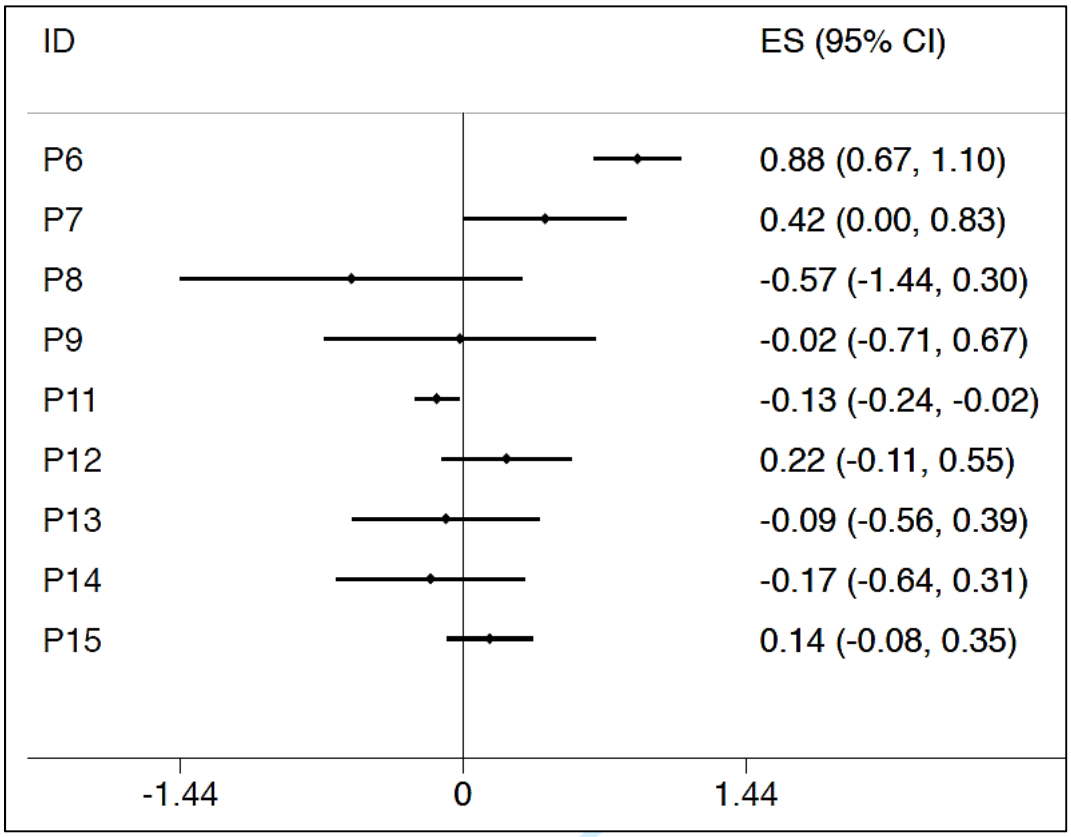
Appendix Figure 5 Legend: Data from this figure was extracted from the study published by Jaeschke et al in 1991, which investigates the effect of amitriptyline or placebo on tender point changes count in patients with fibromyalgia. The average treatment effect is 1.320 (0.404 to 2.236).

Appendix Figure 6: Patients with peptic ulcers, oesophagitis grade I, II, or III, or with reflux or ulcer-like symptom profiles were treated with cimetidine or placebo and its effect on a 6-point symptom scale[4]



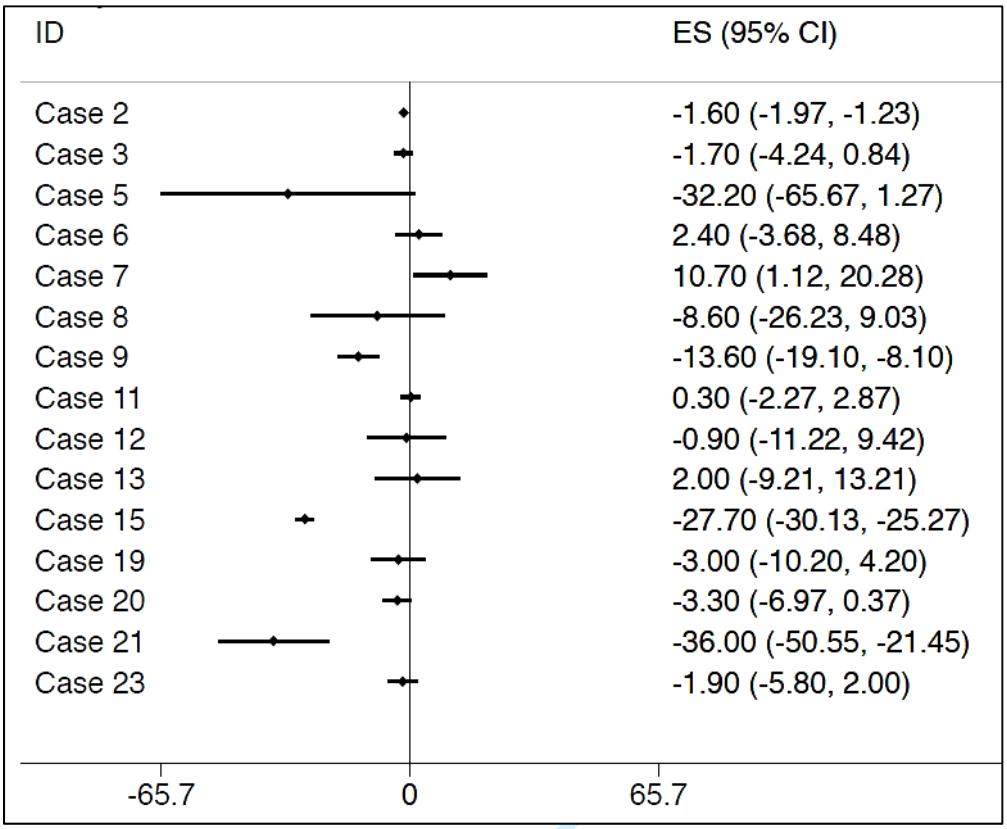
Appendix Figure 6 Legend: Data from this figure was extracted from the study published by Johannessen et al in 1992, which investigates the effect of cimetidine or placebo on a 6-point symptom scale in patients with peptic ulcers, oesophagitis grade I, II, or III, or with reflux or ulcer-like symptom profiles. The average treatment effect is 0.698 (0.466 to 0.931).

Appendix Figure 7: Patients with irreversible chronic airflow limitation treated with theophylline or placebo and its effect on dyspnea[5]



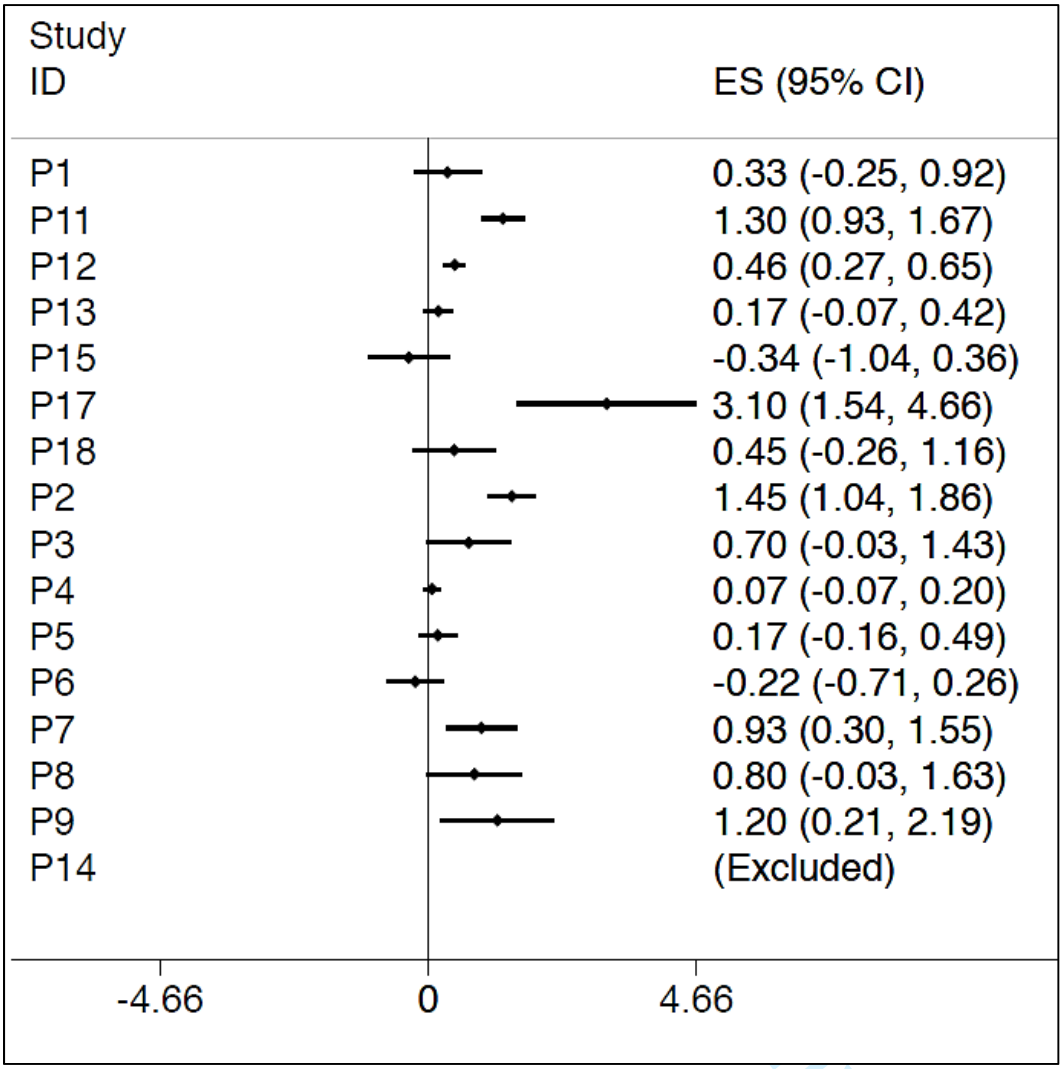
Appendix Figure 7 Legend: Data from this figure was extracted from the study published by Mahon et al in 1996, which investigates the effect of theophylline or placebo on dyspnea in patients with irreversible chronic airflow limitation. The average treatment effect is 0.125 (-0.181 to 0.430).

Appendix Figure 8: Patients with osteoarthritic pain treated with paracetmol and diclofenac and its effect on stiffness[6]



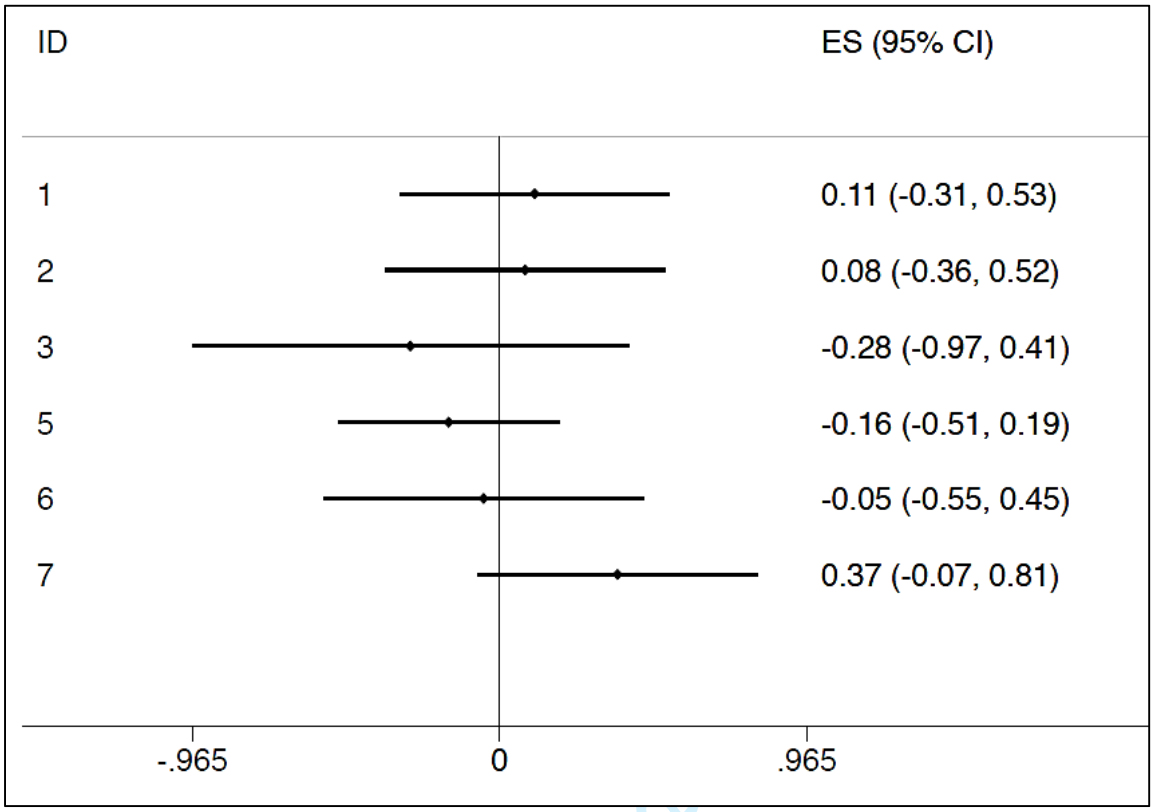
Appendix Figure 8 Legend: Data from this figure was extracted from the study published by March et al in 1994, which investigates the effect of paracetmol and diclofenac on stiffness in patients with osteoarthritic pain. The average treatment effect is mean difference in stiffness (mm).

Appendix Figure 9: Patients with nonreversible chronic airflow limitation treated with either ipratropium bromide, theophylline, salbutamol, or beclomethane (all compared to placebo) and its effect on a 4-item symptom questionnaire[7]



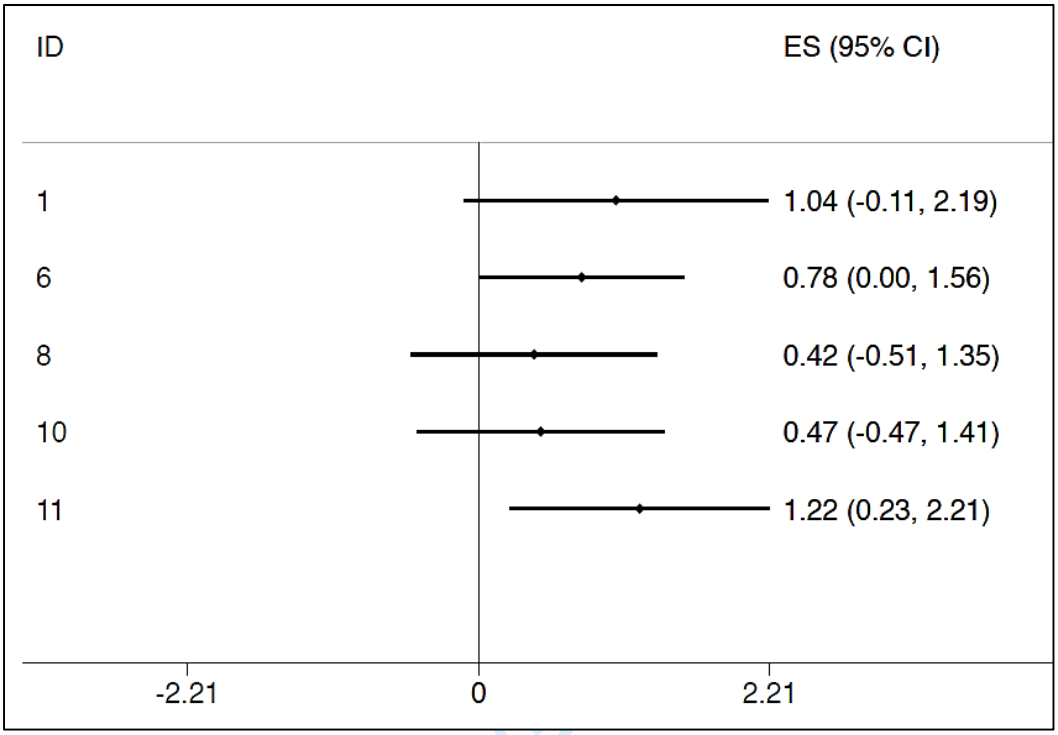
Appendix Figure 9 Legend: Data from this figure was extracted from the study published by Patel et al in 1991, which investigates the effect of ipratropium bromide, theophylline, salbutamol, or beclomethane (all compared to placebo) on a 4-item symptom questionnaire in patients with nonreversible chronic airflow limitation. The average treatment effect is 0.240 (0.131 to 0.350).

Appendix Figure 10: Patients previously taking warfarin for either atrial fibrillation or deep vein thrombosis treated with apo-warfarin and 20coumadin and its effect on international normalized ratio[8]



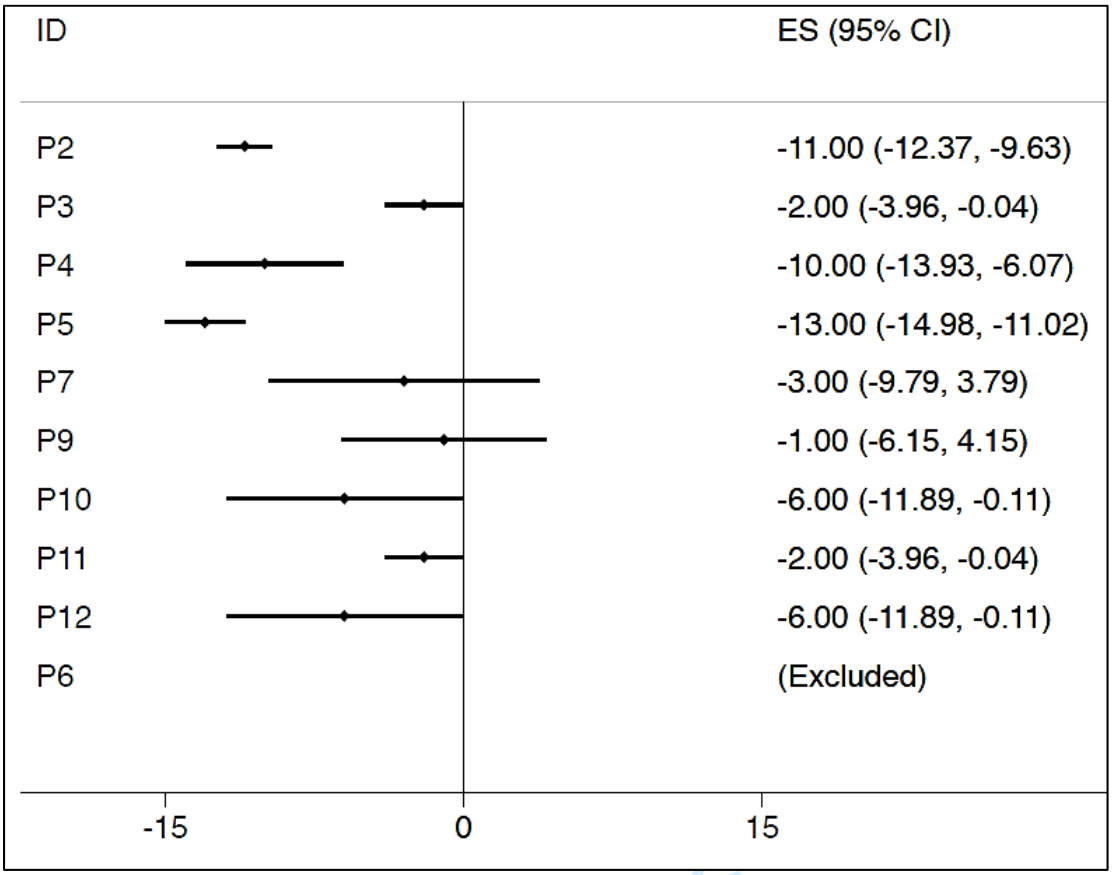
Appendix Figure 10 Legend: Data from this figure was extracted from the study published by Pereira et al in 1995, which investigates the effect of apo-warfarin and Coumadin on international normalized ratio in patients previously taking warfarin for either atrial fibrillation or deep vein thrombosis. The average treatment effect is 0.027 (-0.155 to 0.209).

Appendix Figure 11: Hospitalized children and adolescents with attention-deficit hyperactivity disorder treated with methylphenidate and placebo and its effect on Conners 15-item rating scale scores[9]



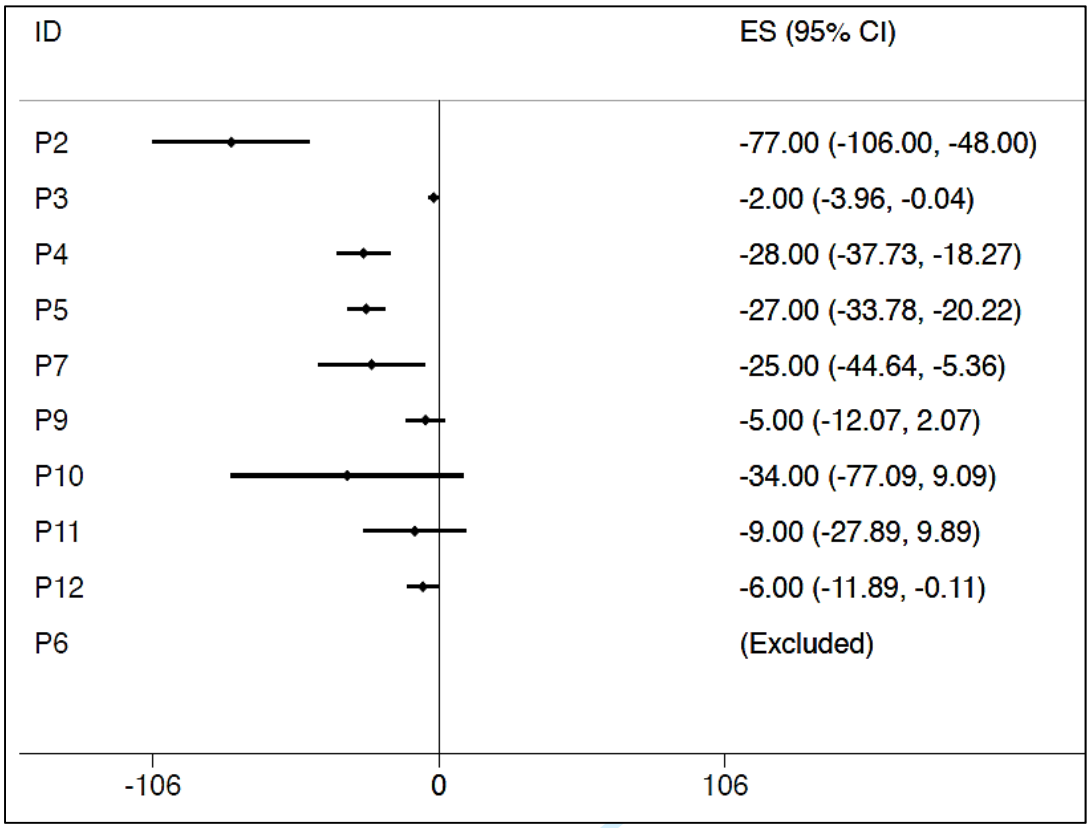
Appendix Figure 11 Legend: Data from this figure was extracted from the study published by Wallace et al in 1994, which investigates the effect of methylphenidate and placebo on Conners 15-item rating scale scores in hospitalized children and adolescents with attention-deficit hyperactivity disorder. The average treatment effect is 0.759 (0.341 to 1.178).

Appendix Figure 12: Patients already prescribed quinine treated with quinine sulphate and placebo, and its effect on changes in number of cramps[10]



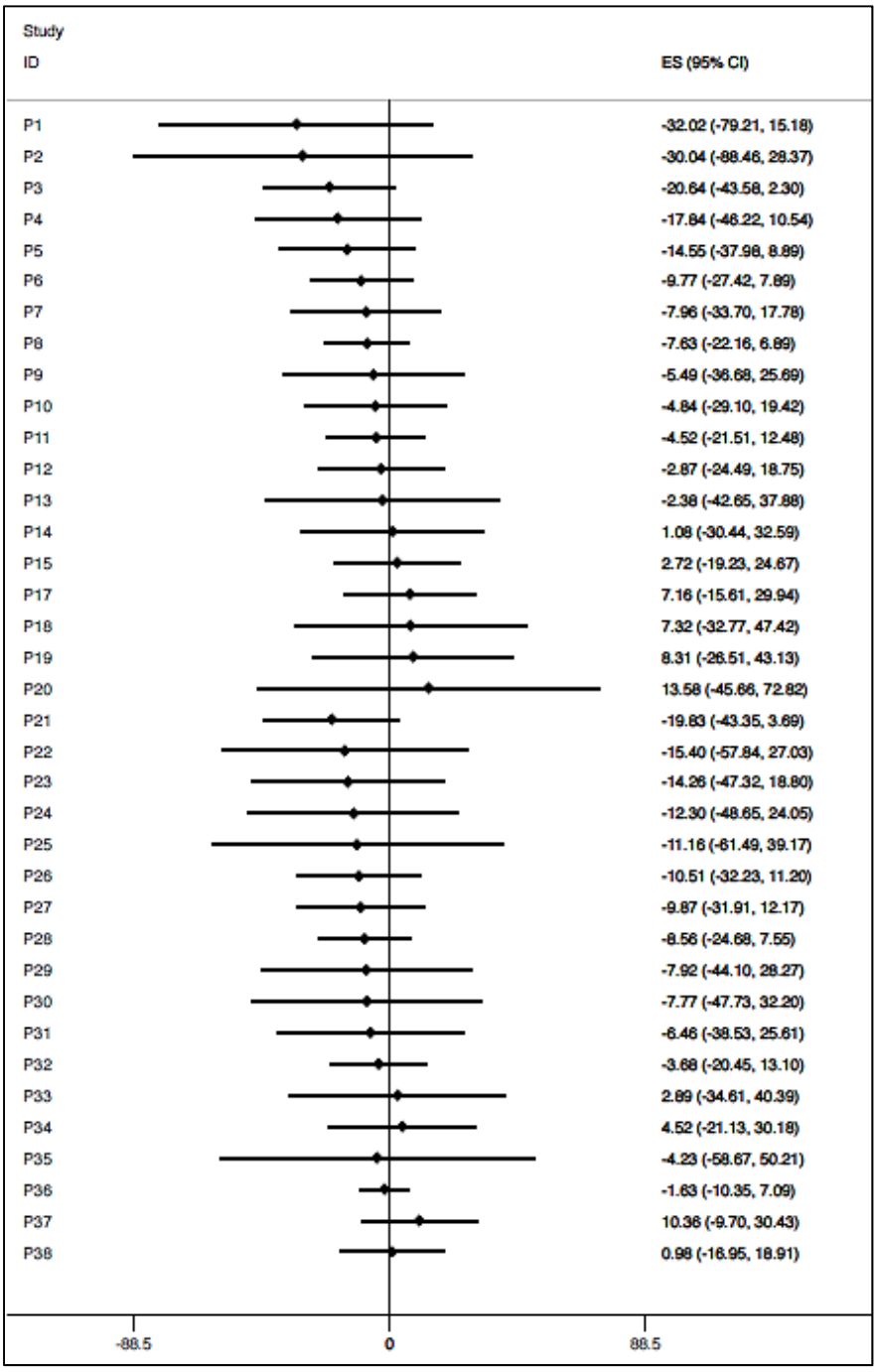
Appendix Figure 12 Legend: Data from this figure was extracted from the study published by Woodfield et al in 2005, which investigates the effect of quinine sulphate and placebo on changes in number of cramps in patients already prescribed quinine. The average treatment effect is -18.823 (-28.527 to -9.120).

Appendix Figure 13: Patients already prescribed quinine treated with quinine sulphate and placebo, and its effect on total days with cramps[10]



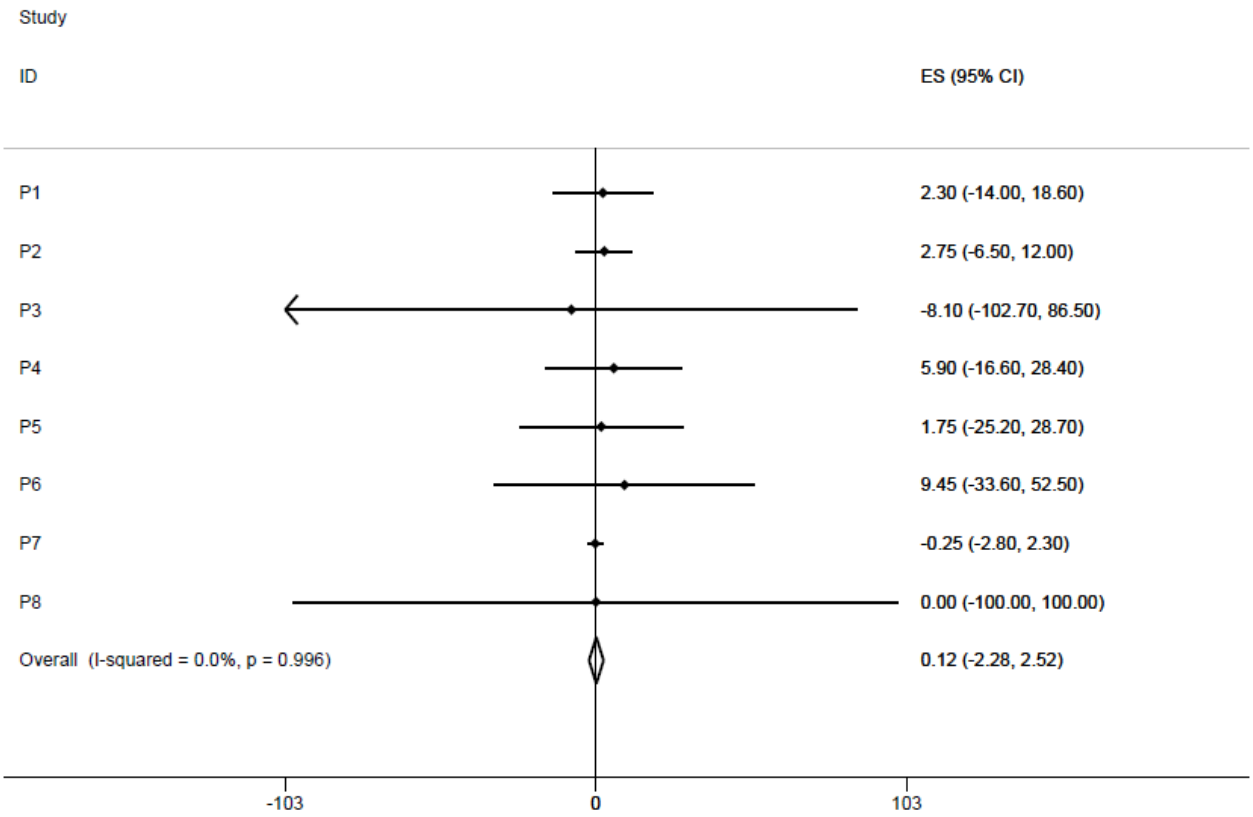
Appendix Figure 13 Legend: Data from this figure was extracted from the study published by Woodfield et al in 2005, which investigates the effect of quinine sulphate and placebo on total days with cramps in patients already prescribed quinine. The average treatment effect is -6.181 (-9.798 to -2.563).

Appendix Figure 14: Patients with fibromyalgia syndrome treated with amitriptyline and the combination amitriptyline and fluoxetine and its effect on the Fibromyalgia Impact Questionnaire[11]



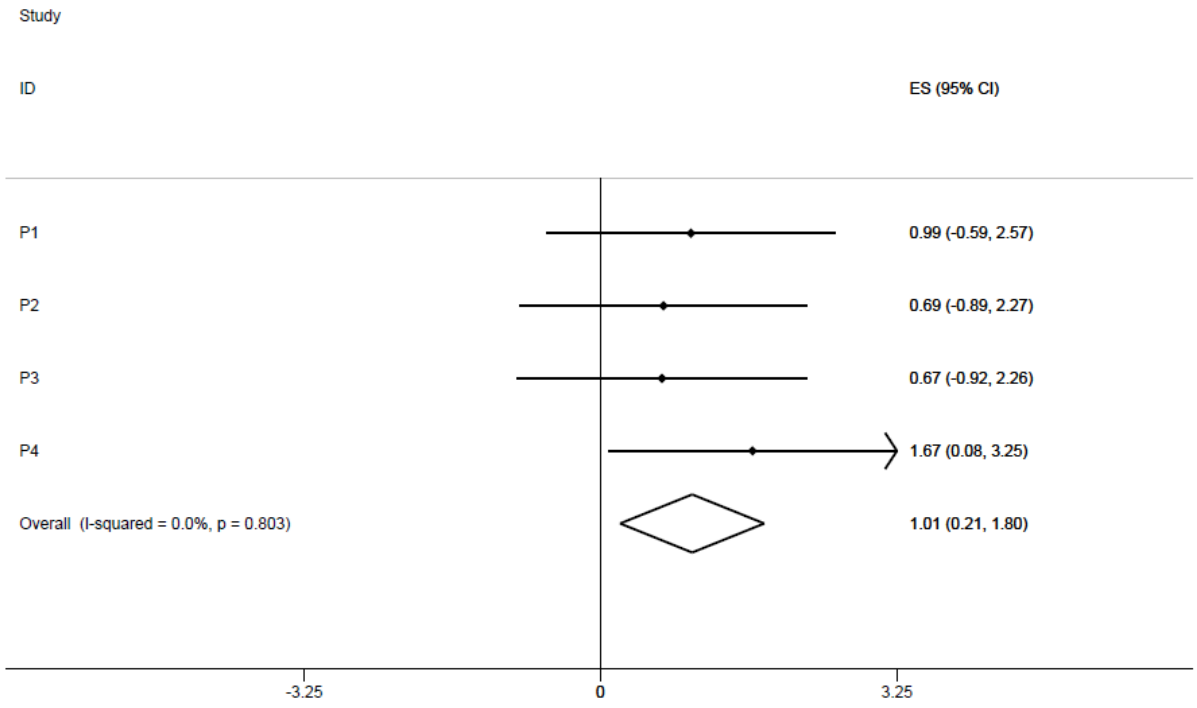
Appendix Figure 14 Legend: Data from this figure was extracted from the study published by Zucker et al in 2006, which investigates the effect of amitriptyline and the combination amitriptyline and fluoxetine on Fibromyalgia Impact Questionnaire in patients with fibromyalgia syndrome. The average treatment effect is -5.019 (-8.784 to -1.254).

Appendix 15: Patients with prior statin-related myalgia with or without mild elevation of creatine kinase levels treated with statin and placebo and its effects on VAS myalgia score[12]



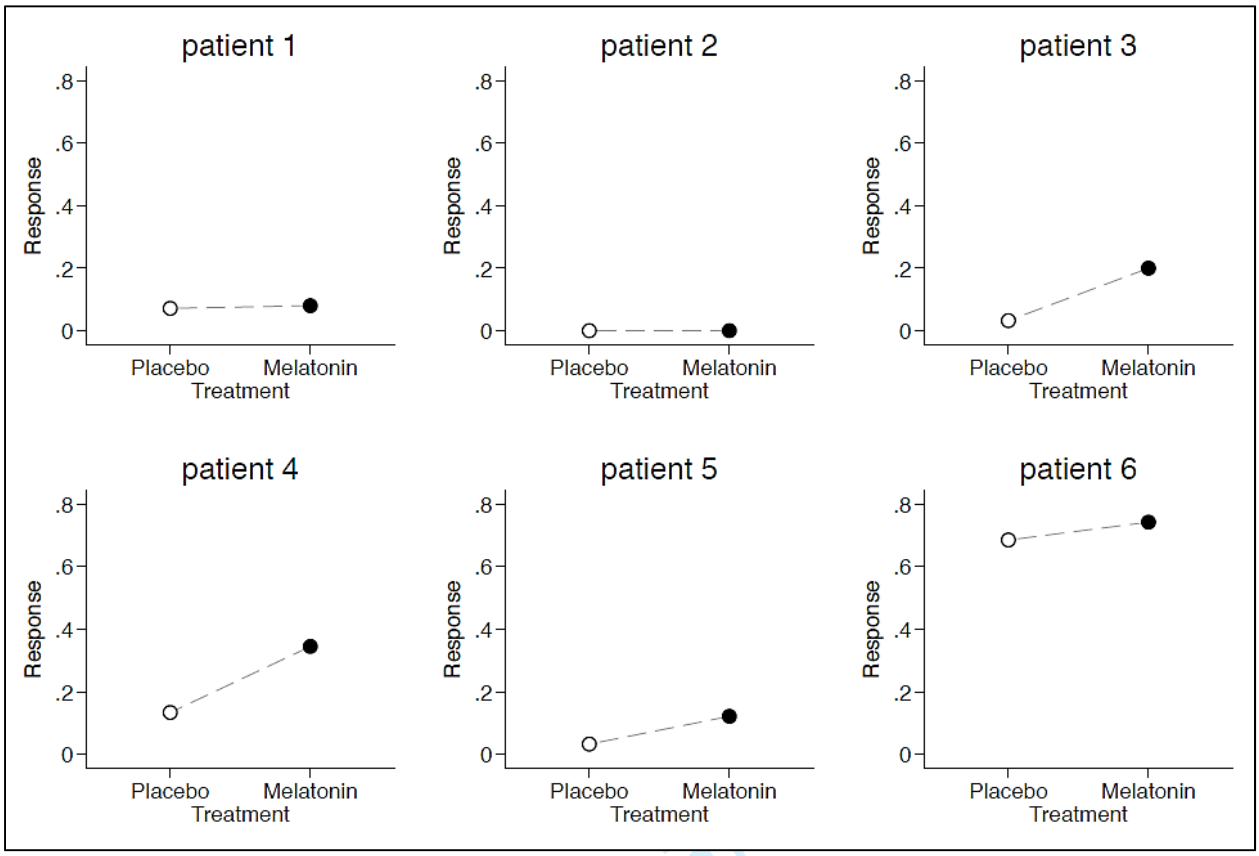
Appendix 15 Figure Legend: Data from this figure was extracted from the study published by Joy et al in 2014, which investigates the effect of statin versus placebo on VAS myalgia score in patients with hyperlipidemia. The average treatment effect is 0.12 (-2.28 to 2.52).

Appendix Figure 16: Patients with myasthenia gravis with acetylcholine receptor antibodies treated with ephinpherin and placebo and its effect on QMG score[13]



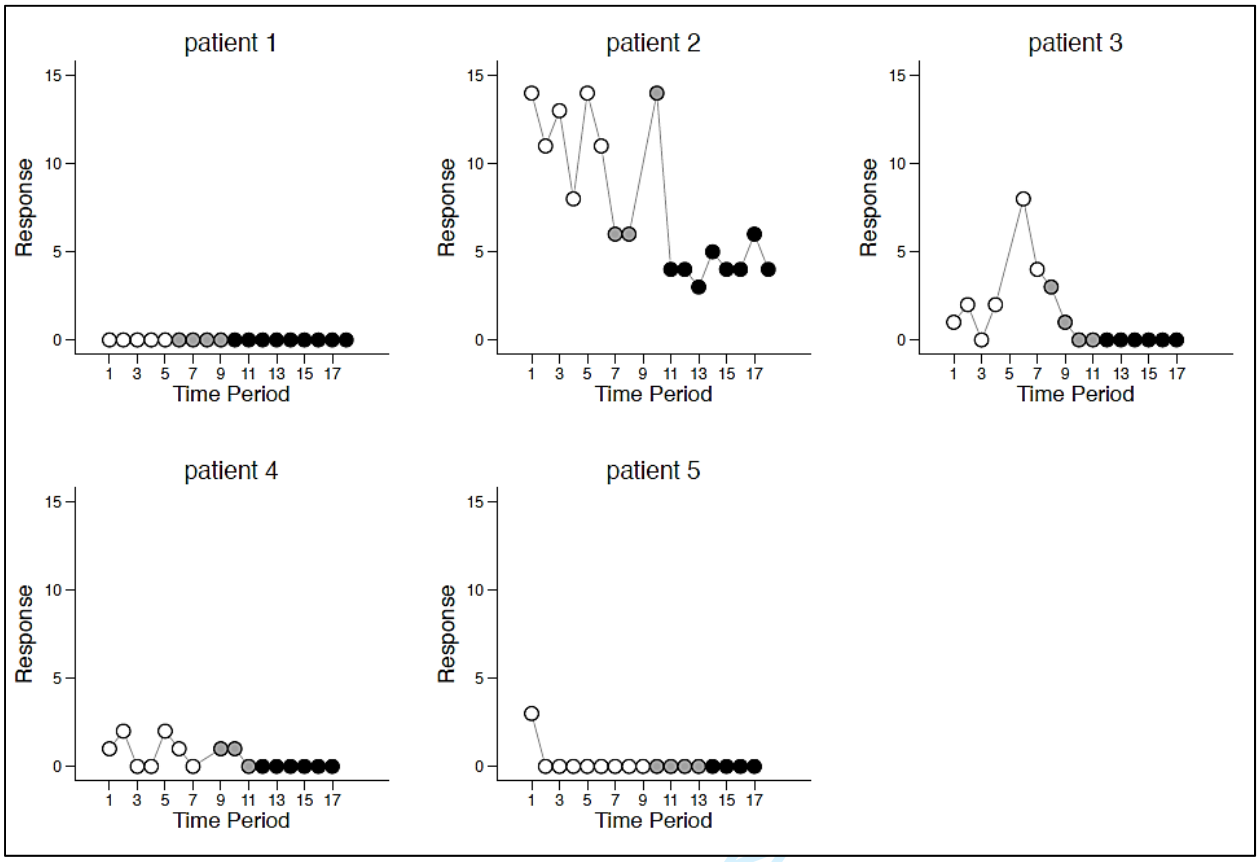
Appendix Figure 16 Legend: Data from this figure was extracted from the study published by Lipkin et al in 2017, which investigates the effect of with ephinpherin and placebo and its effect on QMG score in patients with autoimmune myasthenia gravia. The average treatment effect is 1.01 (0.21 to 1.80).

Appendix Figure 17: Children with mental retardation and fragmented sleep treated with melatonin and placebo and its effect on nights without awakening[14]



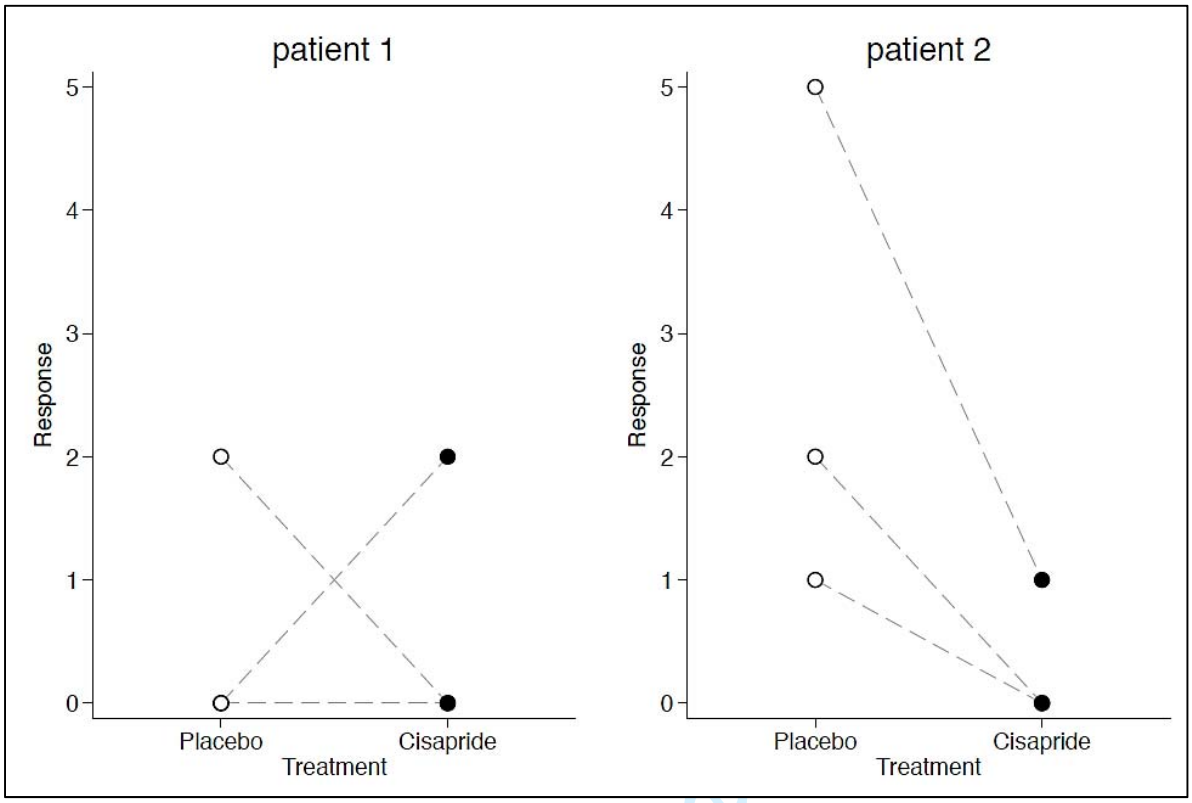
Appendix Figure 17 Legend: Data from this figure was extracted from the study published by Camfield et al in 1996, which investigates the effect of melatonin and placebo on nights without awakening in children with mental retardation and fragmented sleep. The average treatment effect is 0.84 (0.20 to 1.48). White circles indicate placebo; black circles indicate melatonin.

Appendix Figure 18: Patients with traumatic spinal cord lesions treated with baclofen and placebo and its effect on anxiety[15]



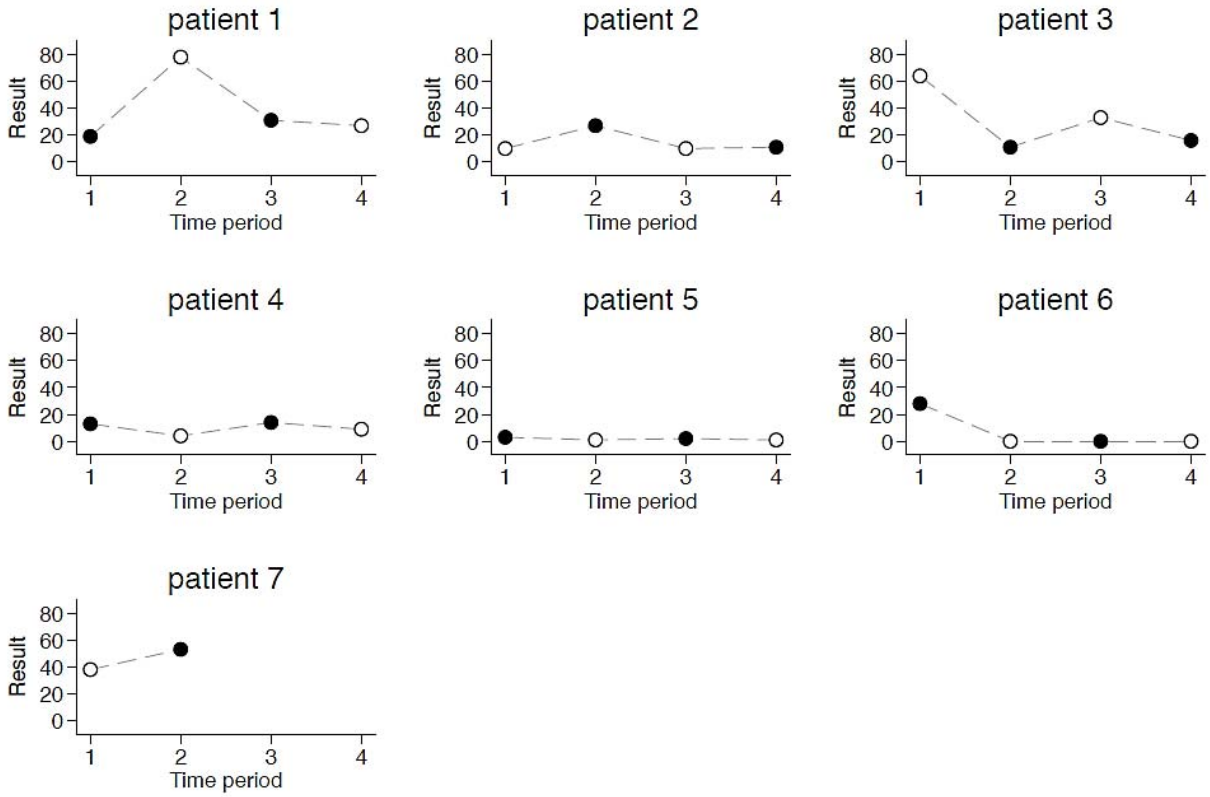
Appendix Figure 18 Legend: Data from this figure was extracted from the study published by Hinderer et al in 1990, which investigates the effect of baclofen and placebo on anxiety in patients with traumatic spinal cord lesions. The average treatment effect is -1.06 (-1.88 to -0.23). White circles indicate placebo; grey circles indicate a half dose (40 mg/day) of baclofen; black circles indicate a full dose (80 mg/day) of baclofen.

Appendix Figure 19: Children with gastroesophageal reflux treated with cisapride and placebo and its effect on emetic episodes per day[16]



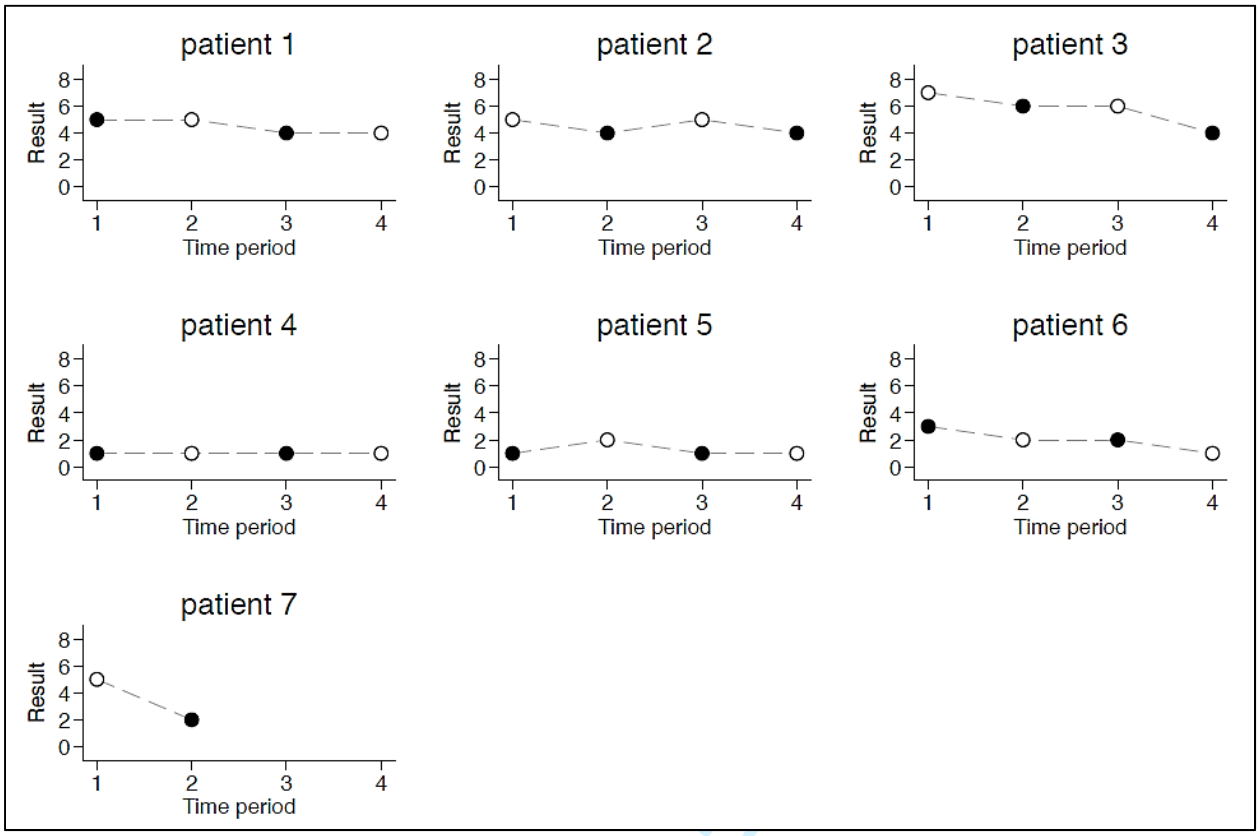
Appendix Figure 19 Legend: Data from this figure was extracted from the study published by Langer et al in 1993, which investigates the effect of cisapride and placebo on emetic episodes per day in children with gastroesophageal reflux. The average treatment effect is -1.20 (-2.49 to 0.09). White circles indicate placebo; black circles indicate cisapride.

Appendix Figure 20: Nonsmokers with ulcerative colitis treated with nicotine gum and placebo and its effect on abdominal pain[17]



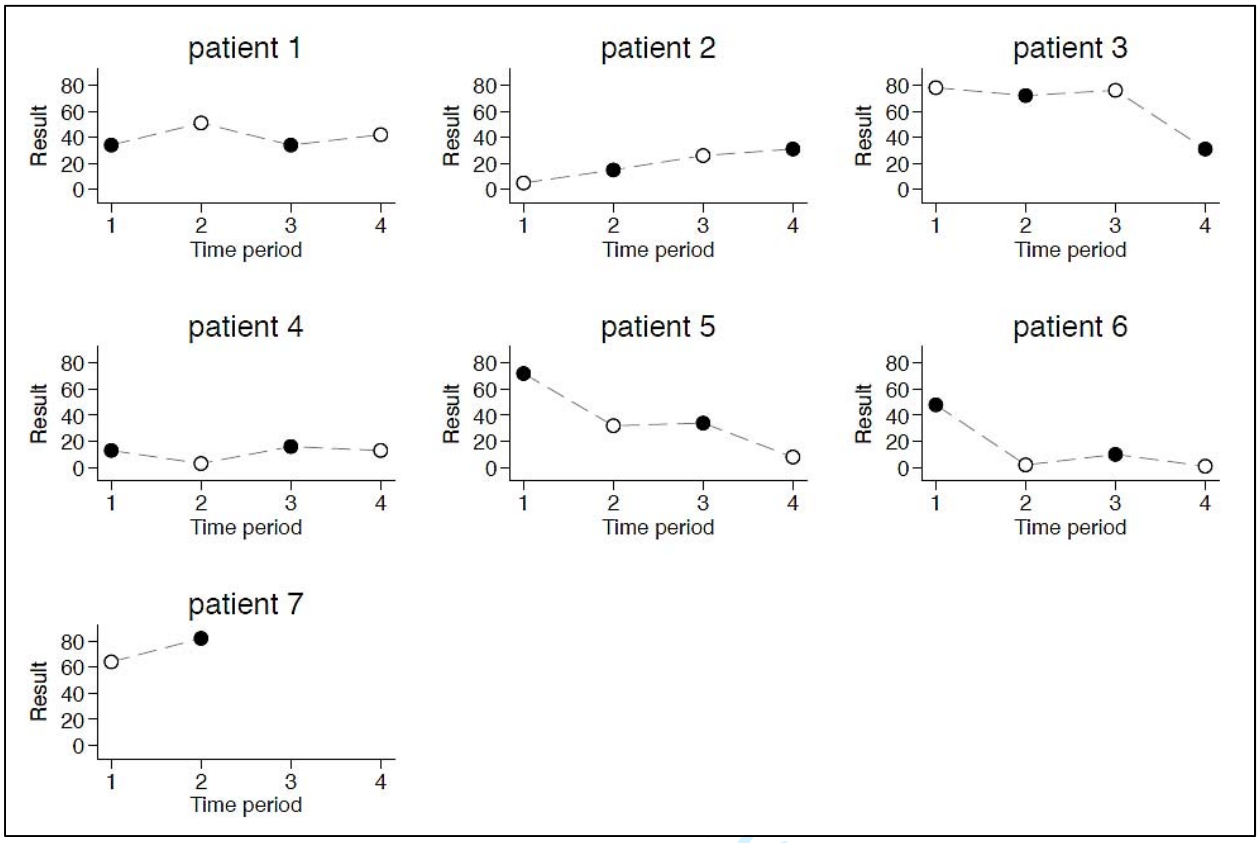
Appendix Figure 20 Legend: Data from this figure was extracted from the study published by Lashner et al in 1990, which investigates the effect of nicotine gum and placebo on abdominal pain in nonsmokers with ulcerative colitis. The average treatment effect is -3.62 (-15.84 to 8.61). White circles indicate placebo gum; black circles indicate nicotine gum.

Appendix Figure 21: Nonsmokers with ulcerative colitis treated with nicotine gum and placebo and its effect on bowel movements per day[17]



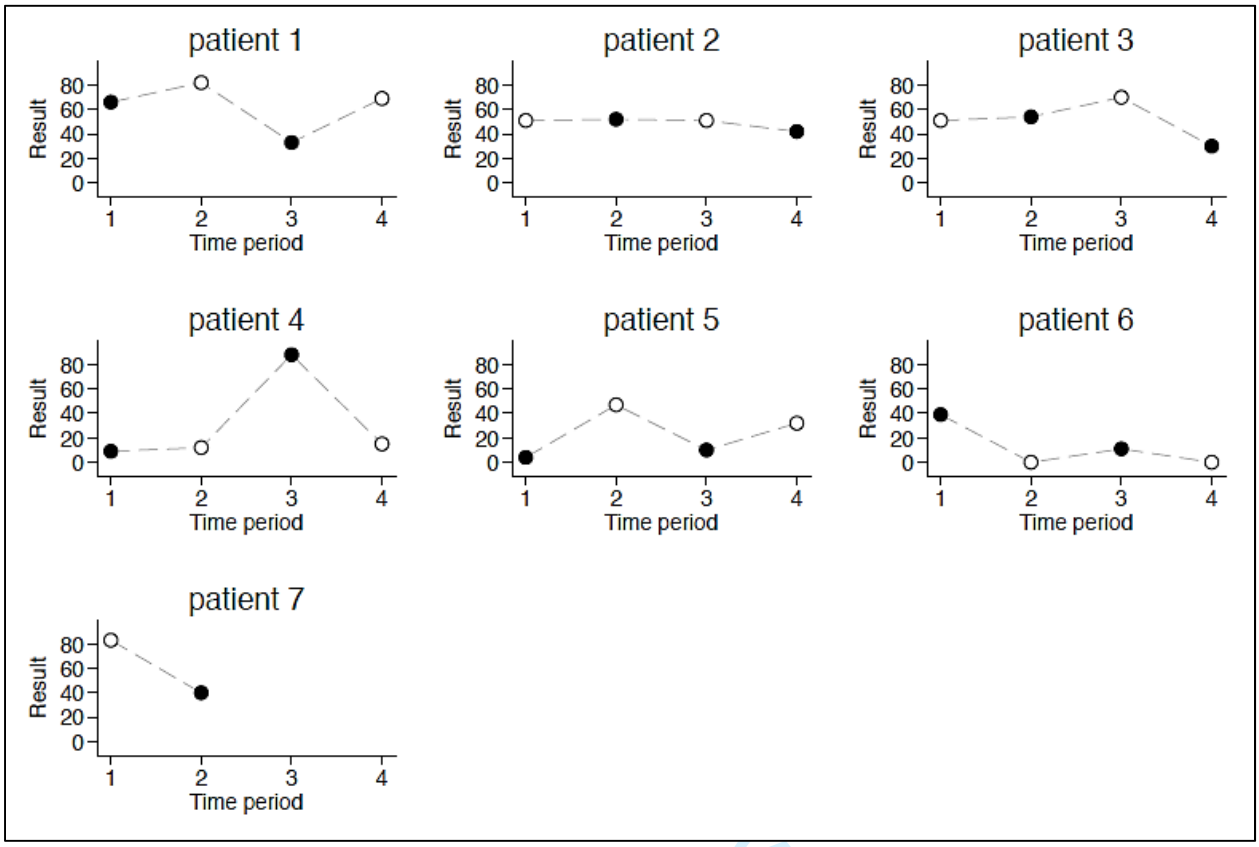
Appendix Figure 21 Legend: Data from this figure was extracted from the study published by Lashner et al in 1990, which investigates the effect of nicotine gum and placebo on bowel movements per day in nonsmokers with ulcerative colitis. The average treatment effect is -0.56 (-1.22 to 0.09). White circles indicate placebo gum; black circles indicate nicotine gum.

Appendix Figure 22: Nonsmokers with ulcerative colitis treated with nicotine gum and placebo and its effect on consistency of bowel movements[17]



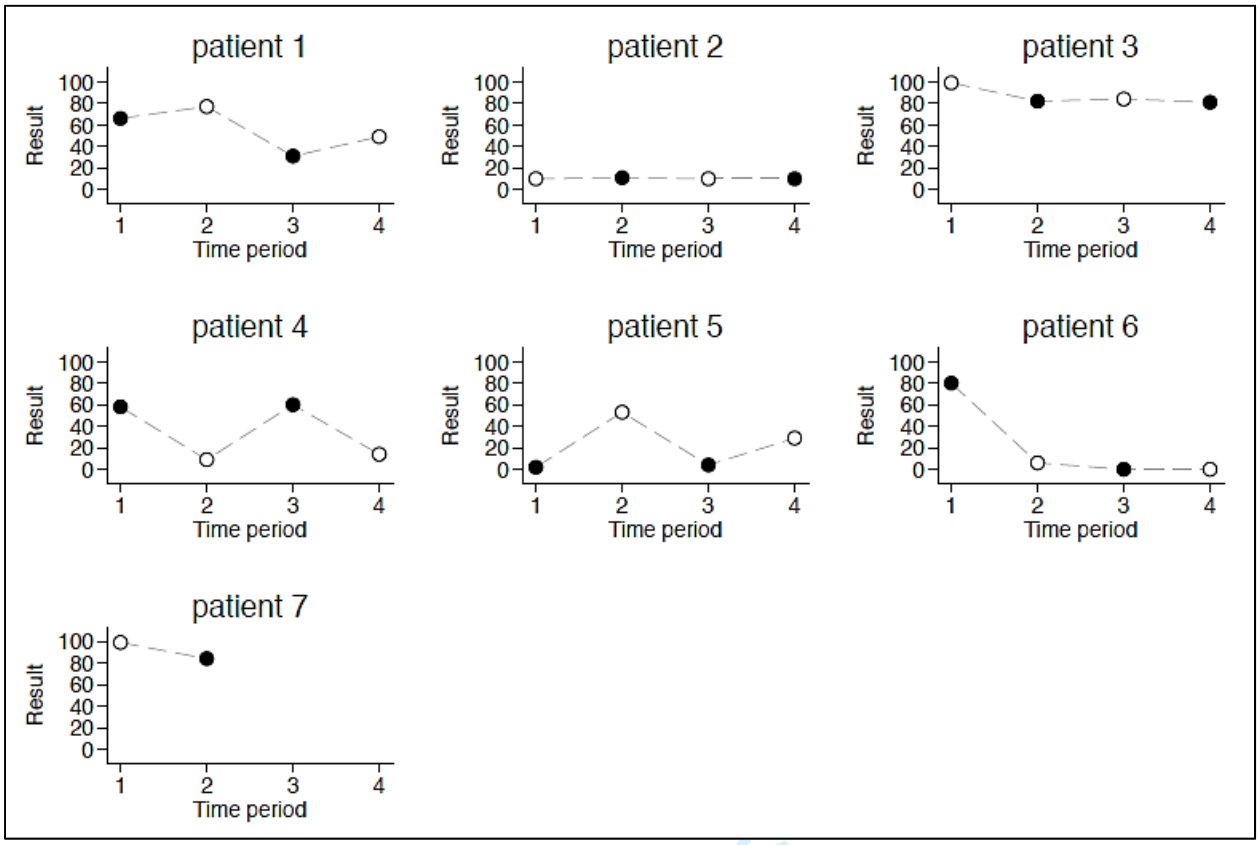
Appendix Figure 22 Legend: Data from this figure was extracted from the study published by Lashner et al in 1990, which investigates the effect of nicotine gum and placebo on consistency of bowel movements in nonsmokers with ulcerative colitis. The average treatment effect is 7.00 (-6.29 to 20.29). White circles indicate placebo gum; black circles indicate nicotine gum.

Appendix Figure 23: Nonsmokers with ulcerative colitis treated with nicotine gum and placebo and its effect on general sense of well-being[17]



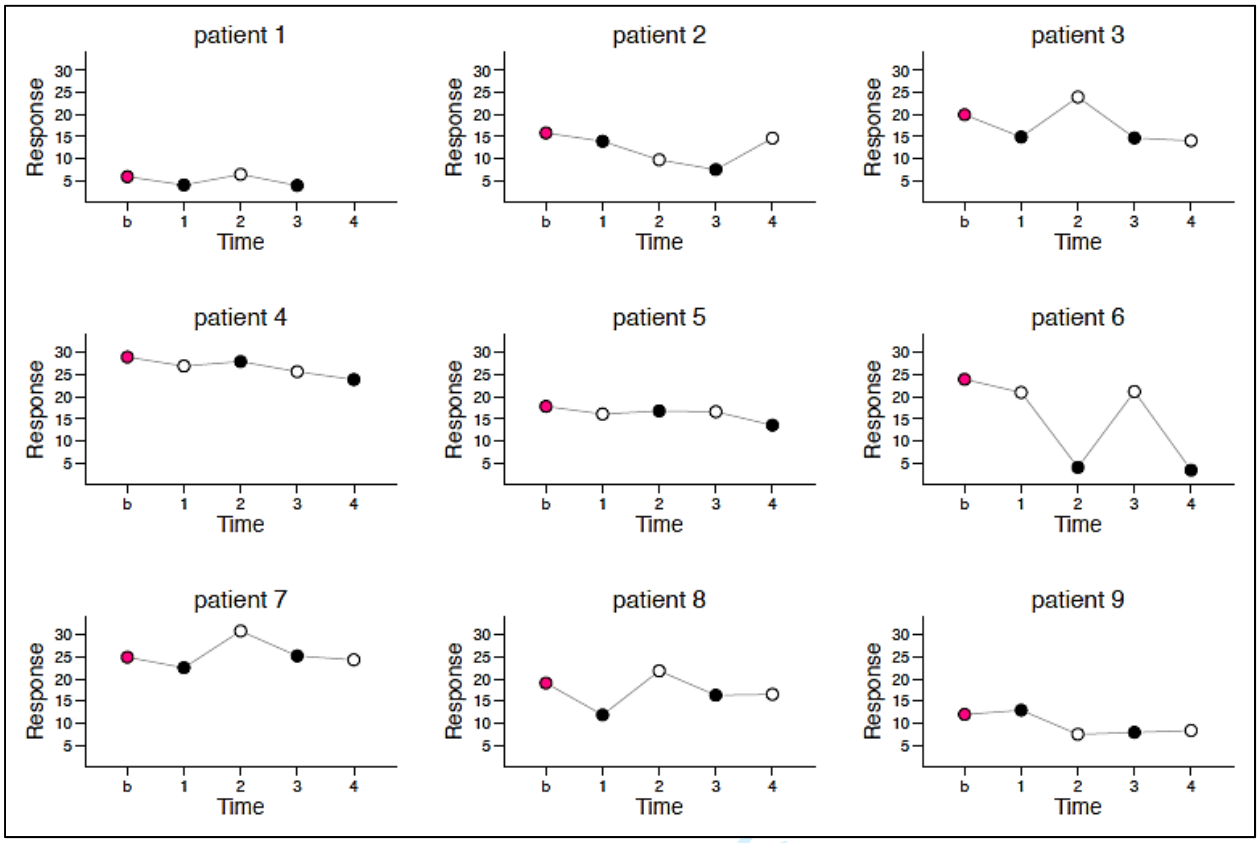
Appendix Figure 23 Legend: Data from this figure was extracted from the study published by Lashner et al in 1990, which investigates the effect of nicotine gum and placebo on general sense of well-being in nonsmokers with ulcerative colitis. The average treatment effect is -6.54 (-23.62 to 10.56). White circles indicate placebo gum; black circles indicate nicotine gum.

Appendix Figure 24: Nonsmokers with ulcerative colitis treated with nicotine gum and placebo and its effect on hematochezia[17]



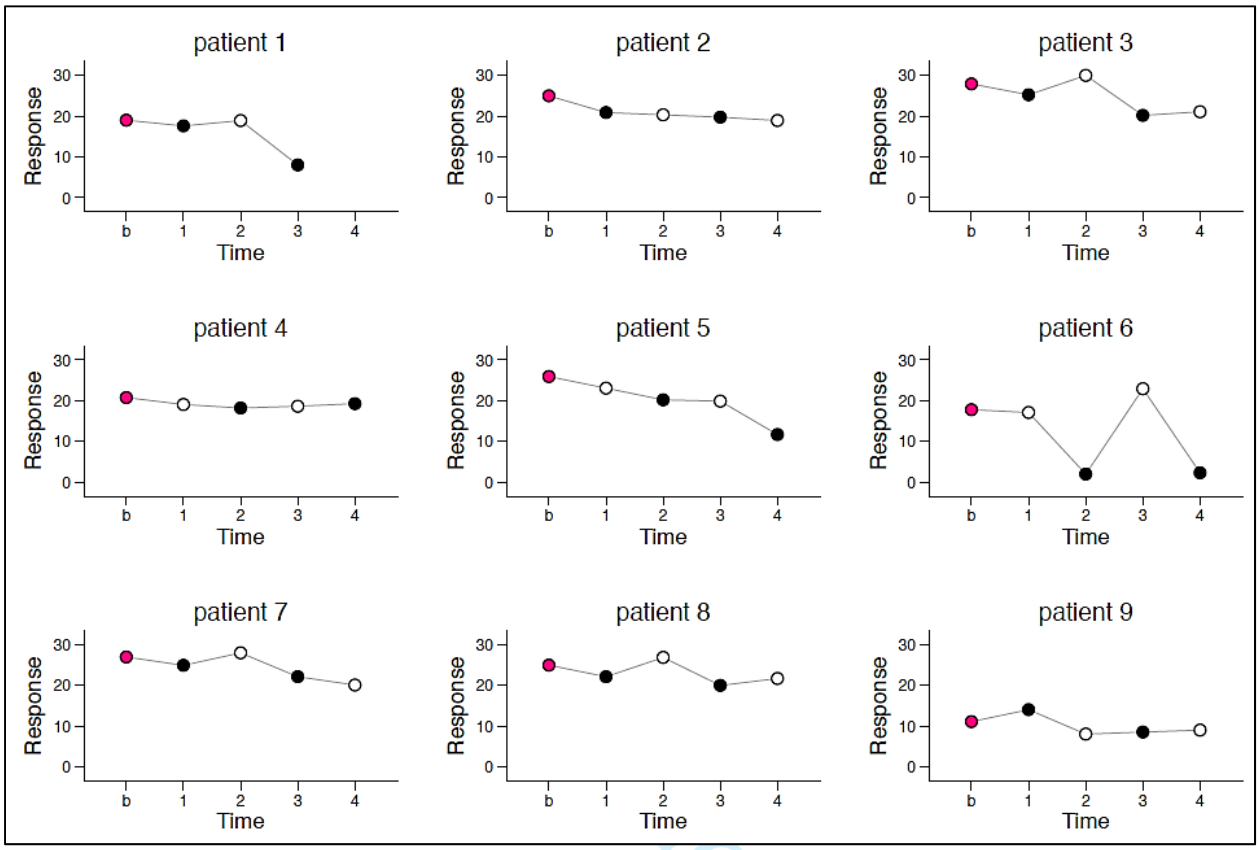
Appendix Figure 24 Legend: Data from this figure was extracted from the study published by Lashner et al in 1990, which investigates the effect of nicotine gum and placebo on hematochezia in nonsmokers with ulcerative colitis. The average treatment effect is 2.35 (-17.21 to 21.90). White circles indicate placebo gum; black circles indicate nicotine gum.

Appendix Figure 25: Patients with chronic depression and a diagnosis of major depression or dysthymia treated with sulpiride and placebo and its effect on anxiety[18]



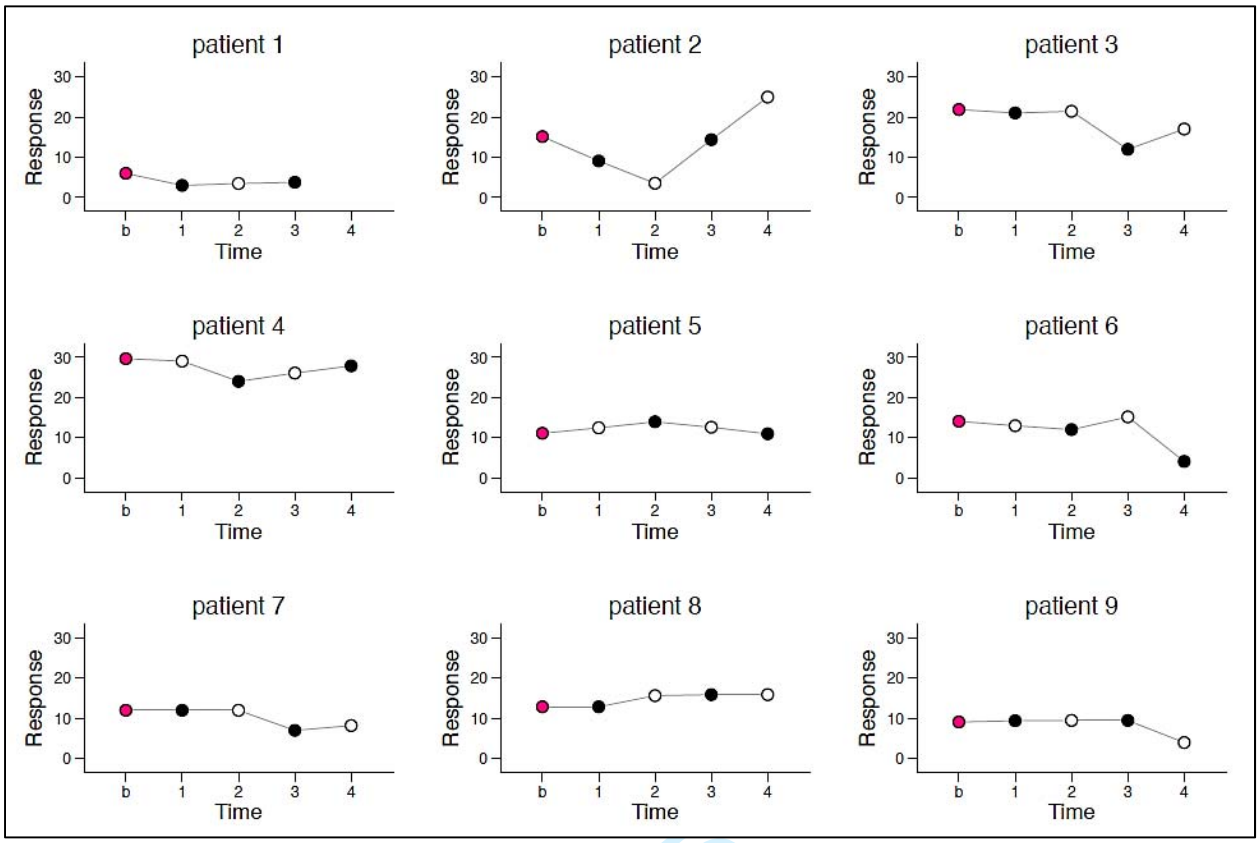
Appendix Figure 25 Legend: Data from this figure was extracted from the study published by Maier et al in 1994, which investigates the effect of sulpiride and placebo on anxiety in patients with chronic depression and a diagnosis of major depression or dysthymia. The average treatment effect is -3.81 (-7.22 to -0.40). Red circles indicate baseline; white circles indicate placebo; black circles indicate sulpiride.

Appendix Figure 26: Patients with chronic depression and a diagnosis of major depression or dysthymia treated with sulpiride and placebo and its effect on depressed mood[18]



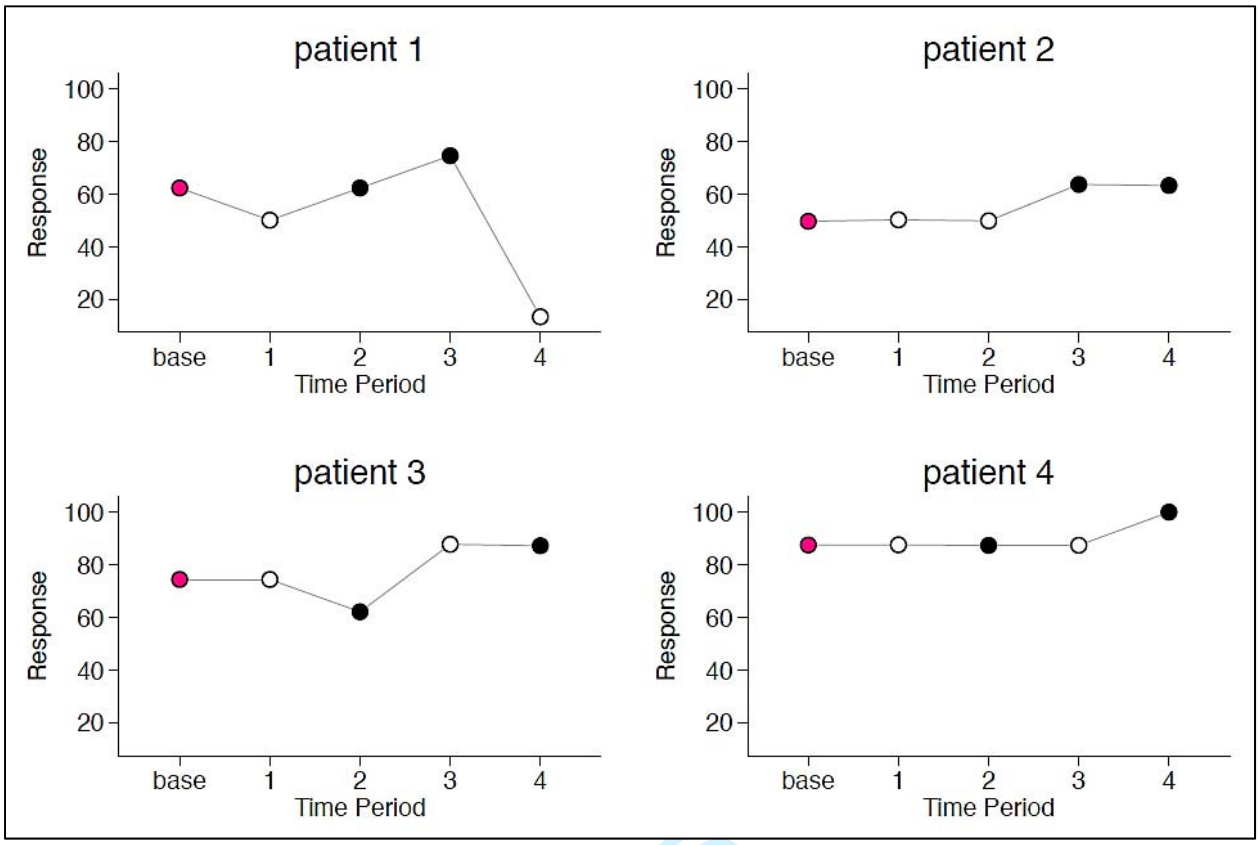
Appendix Figure 26 Legend: Data from this figure was extracted from the study published by Maier et al in 1994, which investigates the effect of sulpiride and placebo on depressed mood in patients with chronic depression and a diagnosis of major depression or dysthymia. The average treatment effect is -3.63 (-7.40 to 0.15). Red circles indicate baseline; white circles indicate placebo; black circles indicate sulpiride.

Appendix Figure 27: Patients with chronic depression and a diagnosis of major depression or dysthymia treated with sulpiride and placebo and its effect on somatization[18]



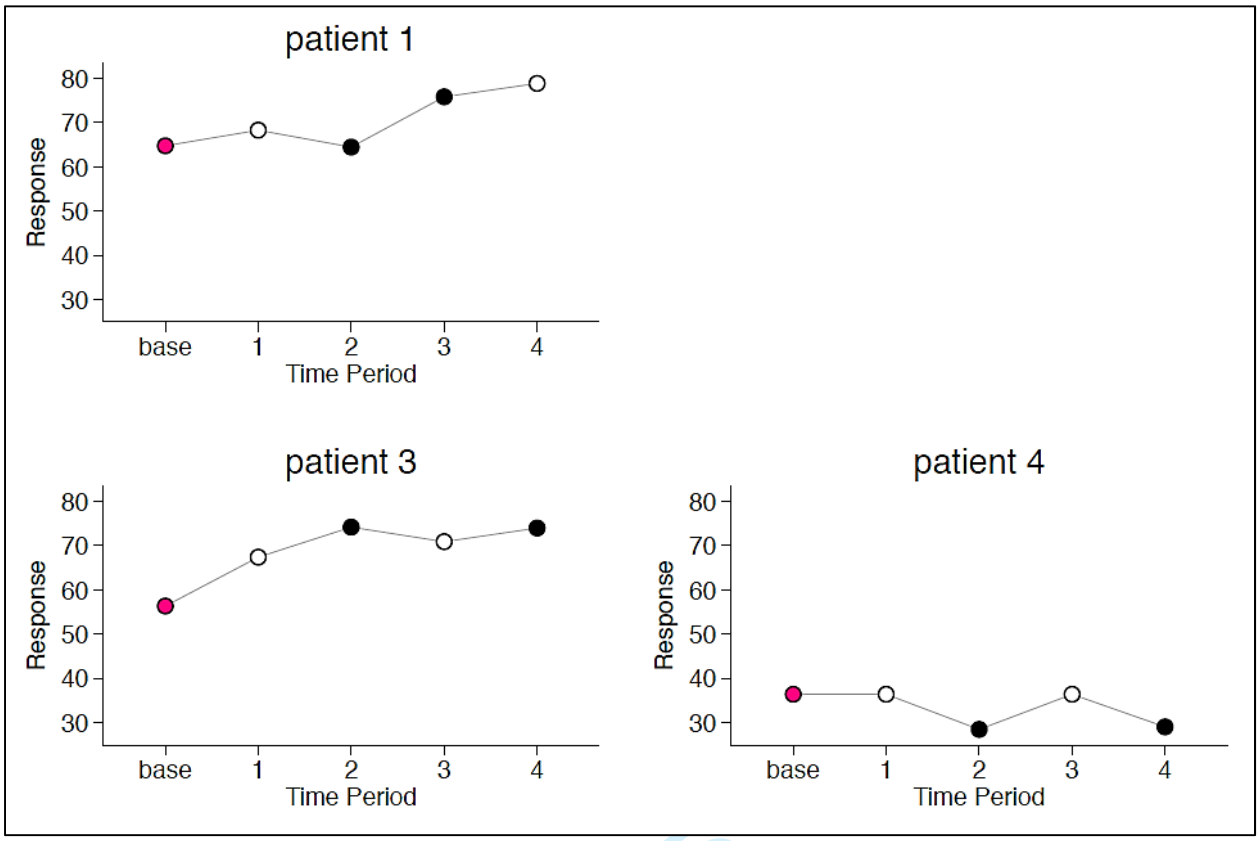
Appendix Figure 27 Legend: Data from this figure was extracted from the study published by Maier et al in 1994, which investigates the effect of sulpiride and placebo on somatization in patients with chronic depression and a diagnosis of major depression or dysthymia. The average treatment effect is -1.50 (-4.20 to 1.21). Red circles indicate baseline; white circles indicate placebo; black circles indicate sulpiride.

Appendix Figure 28: Patients with ataxia from traumatic brain injury treated with ondansetron and placebo and its effect on lower extremity ataxia[19]



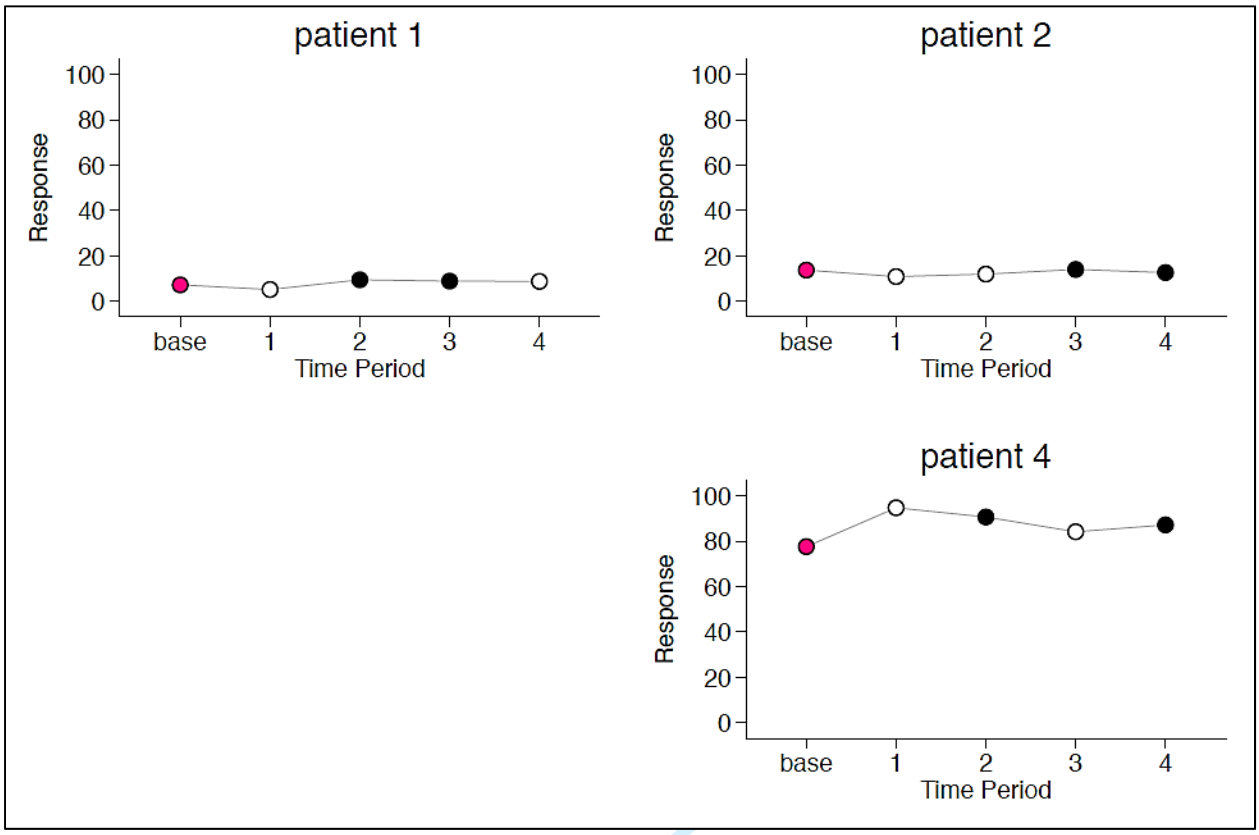
Appendix Figure 28 Legend: Data from this figure was extracted from the study published by Mandelcorn et al in 2004, which investigates the effect of ondansetron and placebo on lower extremity ataxia in patients with ataxia from traumatic brain injury. Each patient received the same treatment. The average treatment effect is 12.49 (-0.85 to 25.84). Red circles indicate baseline; white circles indicate placebo; black circles indicate ondansetron.

Appendix Figure 29: Patients with ataxia from traumatic brain injury treated with ondansetron and placebo and its effect on self-assessment score[19]



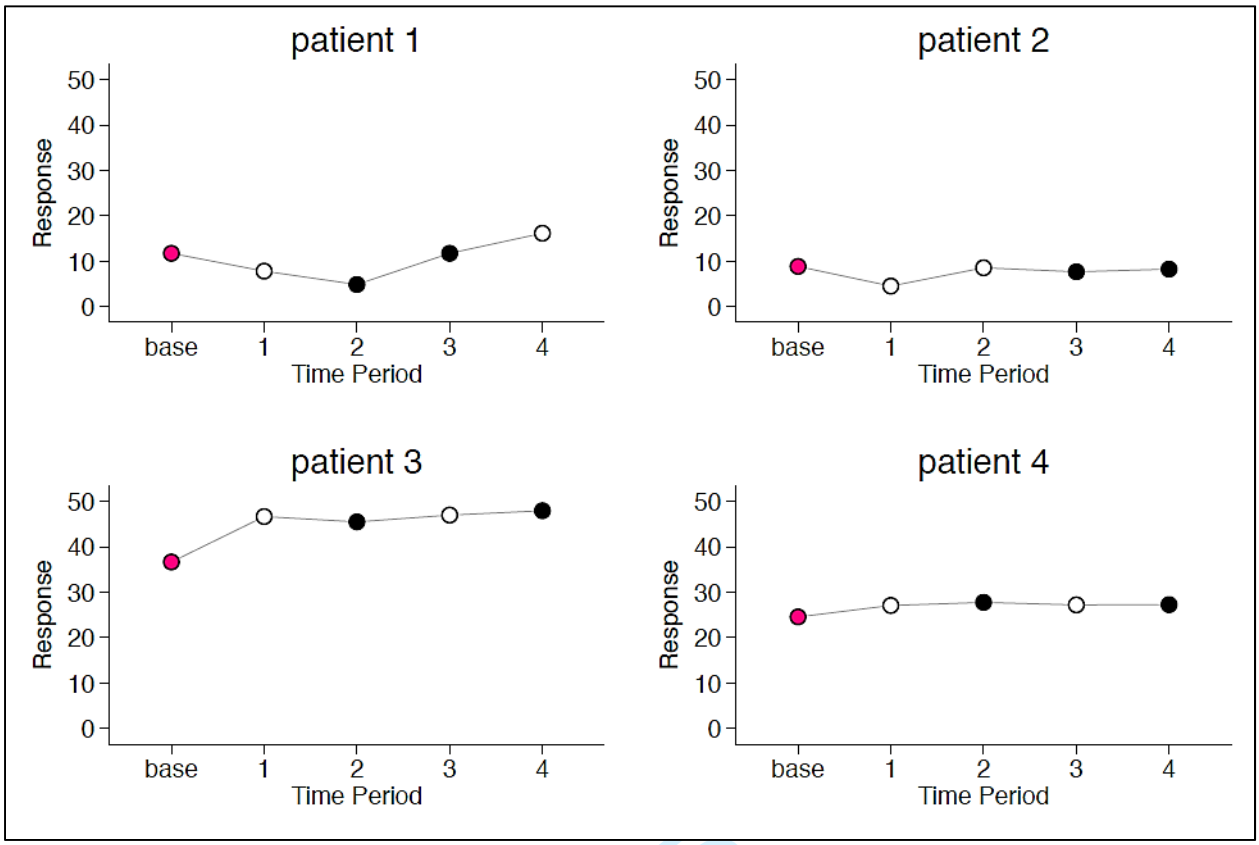
Appendix Figure 29 Legend: Data from this figure was extracted from the study published by Mandelcorn et al in 2004, which investigates the effect of ondansetron and placebo on self-assessment score in patients with ataxia from traumatic brain injury. The average treatment effect is -2.05 (-8.43 to 4.33). Red circles indicate baseline; white circles indicate placebo; black circles indicate ondansetron.

Appendix Figure 30: Patients with ataxia from traumatic brain injury treated with ondansetron and placebo and its effect on truncal ataxia[19]



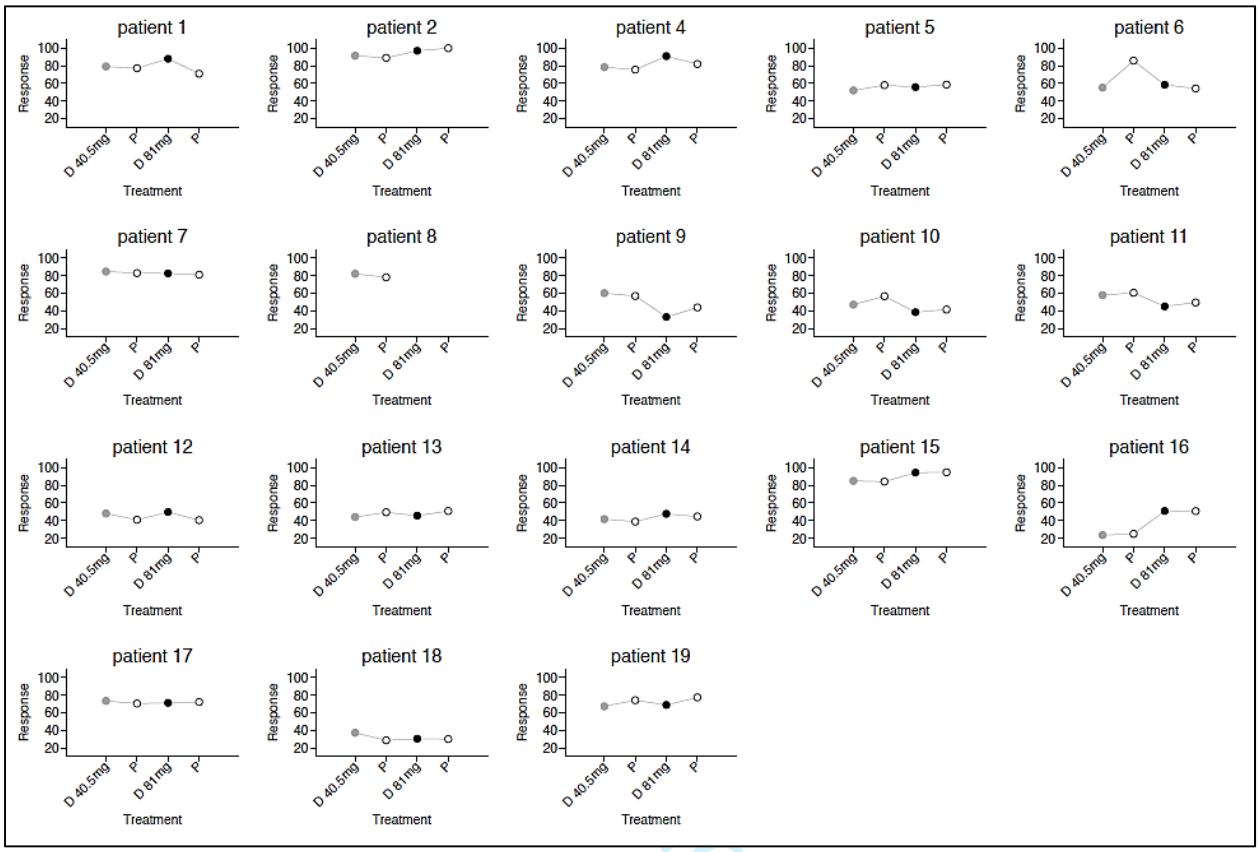
Appendix Figure 30 Legend: Data from this figure was extracted from the study published by Mandelcorn et al in 2004, which investigates the effect of ondansetron and placebo on truncal ataxia in patients with ataxia from traumatic brain injury. The average treatment effect is 1.20 (-2.06 to 4.45). Red circles indicate baseline; white circles indicate placebo; black circles indicate ondansetron.

Appendix Figure 31: Patients with ataxia from traumatic brain injury treated with ondansetron and placebo and its effect on upper extremity ataxia[19]



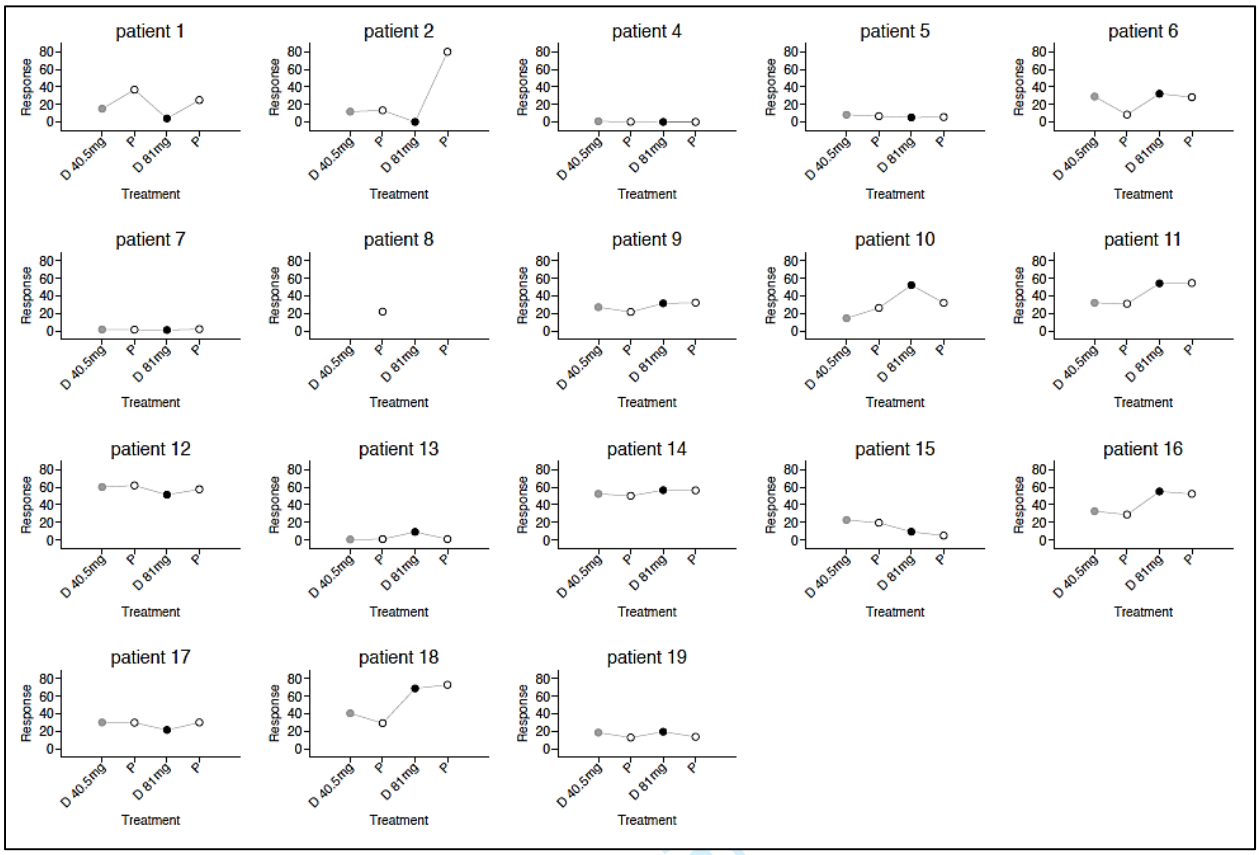
Appendix Figure 31 Legend: Data from this figure was extracted from the study published by Mandelcorn et al in 2004, which investigates the effect of ondansetron and placebo on upper extremity ataxia in patients with ataxia from traumatic brain injury. The average treatment effect is -0.50 (-3.10 to 2.10). Red circles indicate baseline; white circles indicate placebo; black circles indicate ondansetron.

Appendix Figure 32: Patients with chronic neuropathic pain treated with oral dextromethorphan and placebo and its effect on VAS pain intensity[20]



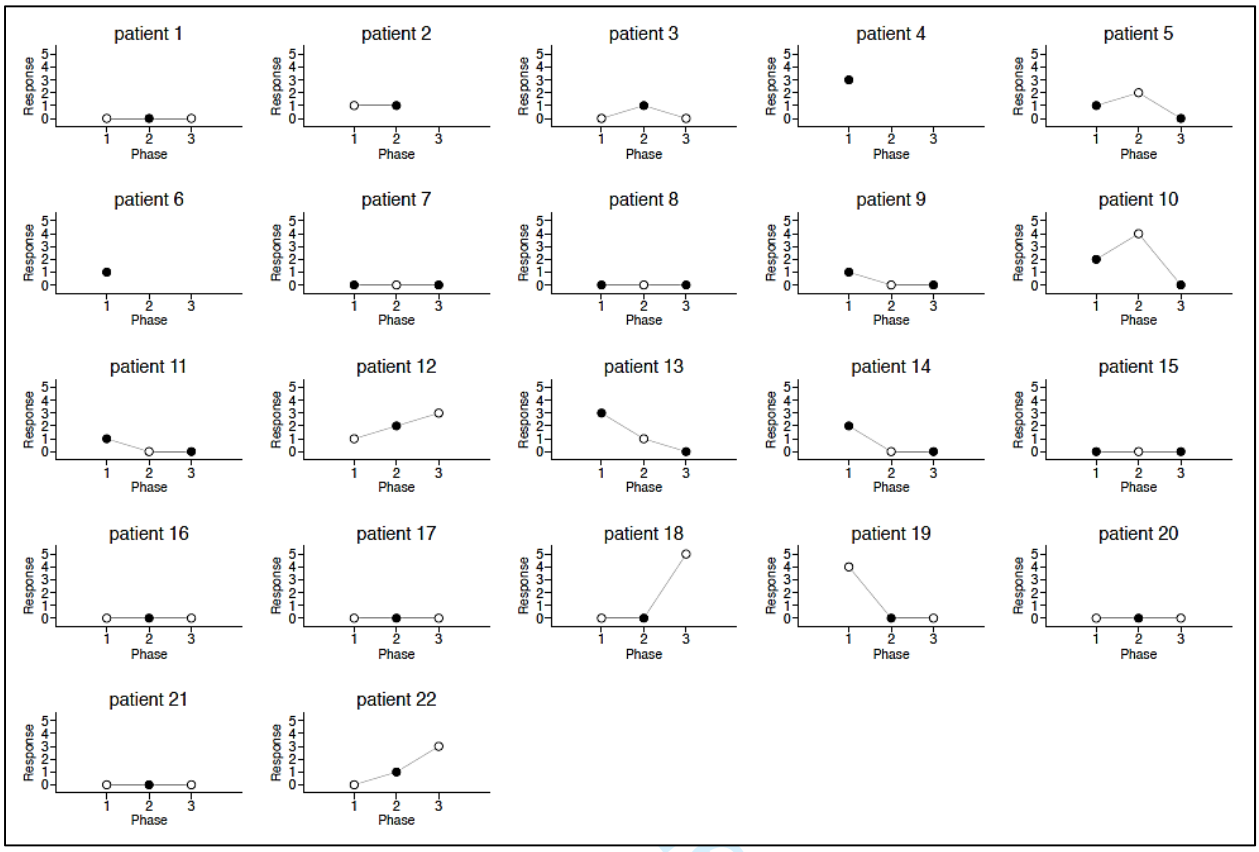
Appendix Figure 32 Legend: Data from this figure was extracted from the study published by McQuay et al in 1994, which investigates the effect of oral dextromethorphan and placebo on VAS pain intensity in patients with chronic neuropathic pain. The average treatment effect is -1.06 (-5.16 to 3.04). Grey circles indicate dextromethorphan 40.5 mg daily; black circles indicate dextromethorphan 81 mg daily; white circles indicate placebo.

Appendix Figure 33: Patients with chronic neuropathic pain treated with oral dextromethorphan and placebo and its effect on VAS relief intensity[20]



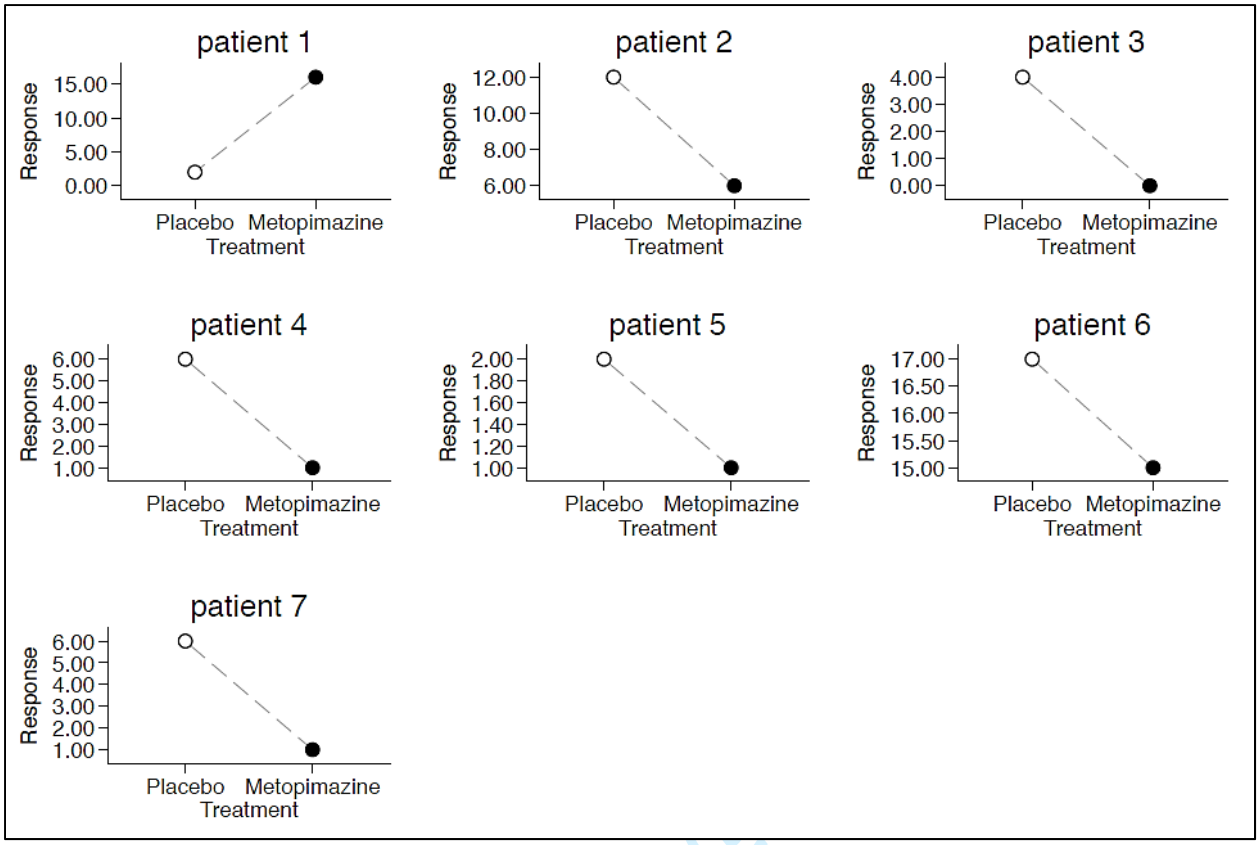
Appendix Figure 33 Legend: Data from this figure was extracted from the study published by McQuay et al in 1994, which investigates the effect of oral dextromethorphan and placebo on VAS relief intensity in patients with chronic neuropathic pain. The average treatment effect is -3.86 (-11.11 to 3.40). Grey circles indicate dextromethorphan 40.5 mg daily; black circles indicate dextromethorphan 81 mg daily; white circles indicate placebo.

Appendix Figure 34: Patients with unstable angina at rest treated with continuous and intermittent injection of isosorbide dinitrate and its effect on incidence of angina[21]



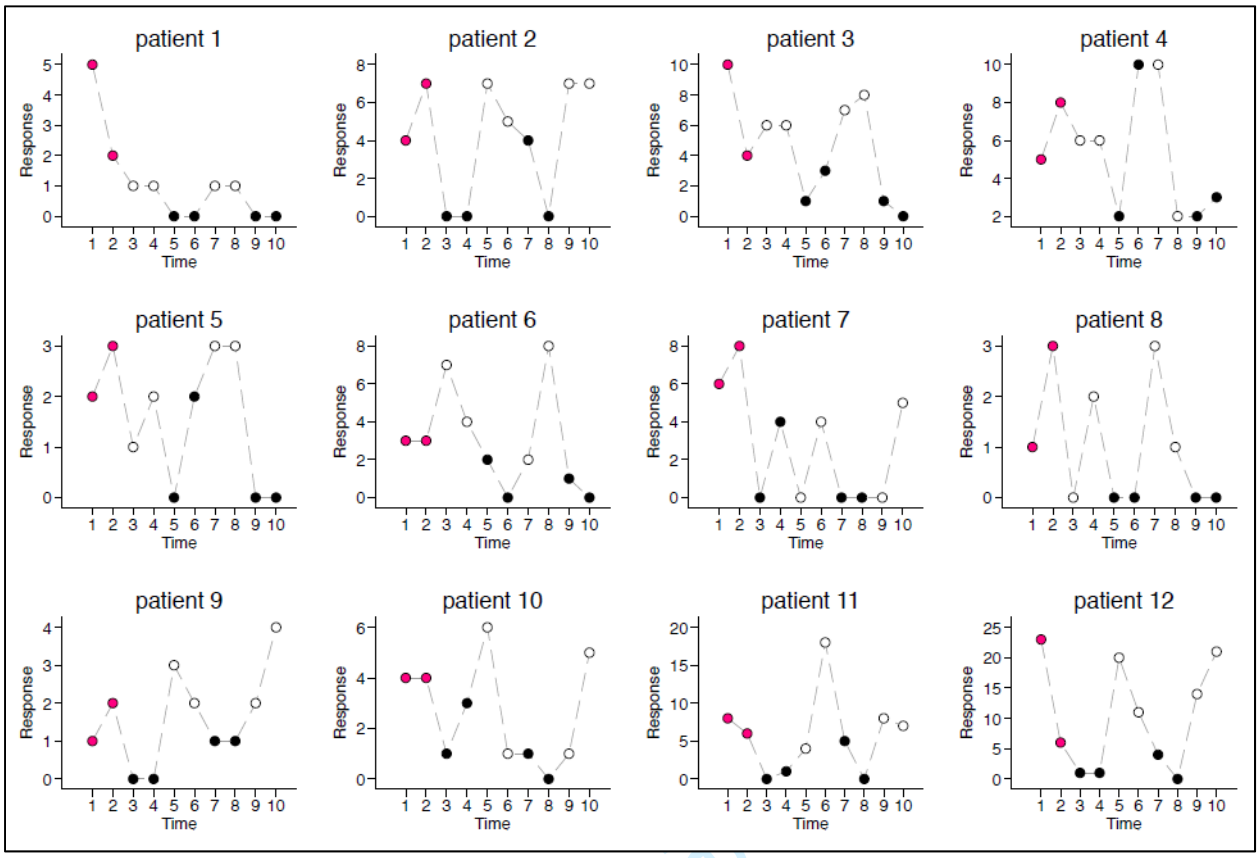
Appendix Figure 34 Legend: Data from this figure was extracted from the study published by Miyazaki et al in 1995, which investigates the effect of continuous and intermittent injection of isosorbide dinitrate on incidence of angina in patients with unstable angina. The average treatment effect is 0.47 (-0.32 to 1.26). White circles indicate continuous injection; black circles indicate intermittent injection.

Appendix Figure 35: Children with brain tumors receiving highly emetogenic therapy treated with ondansetron/metopimazine and ondansetron monotherapy and its effect on emetic episodes per day[22]



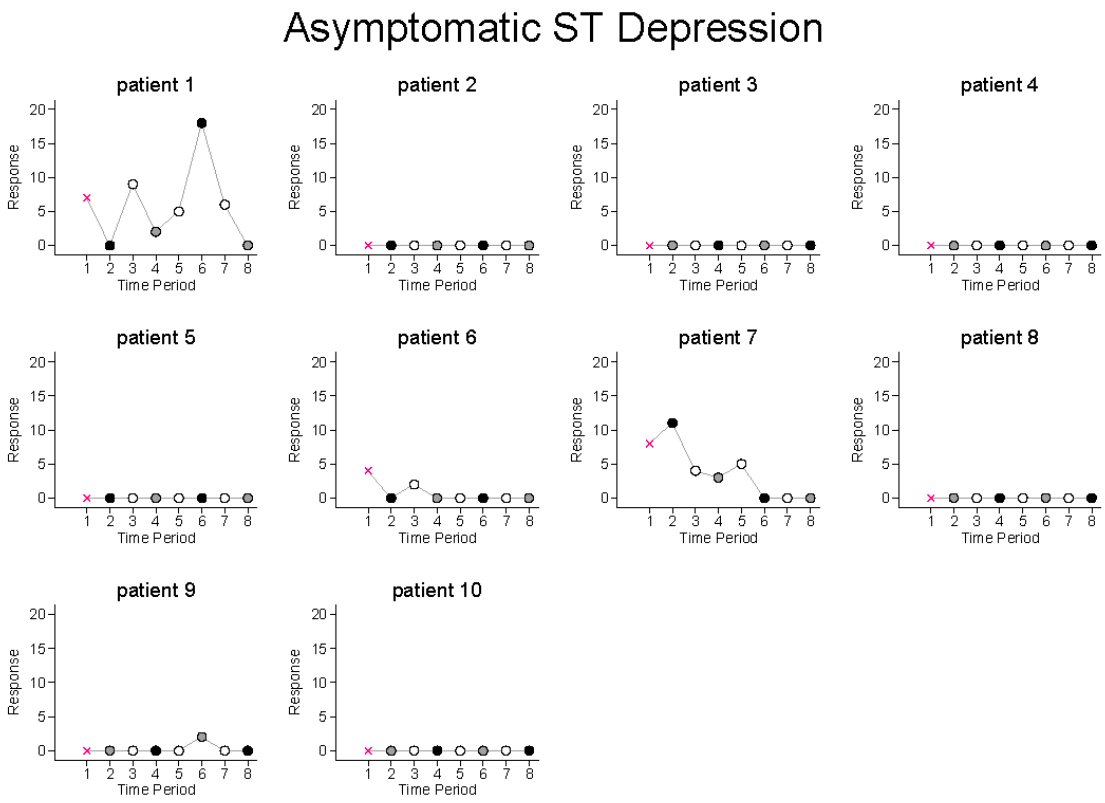
Appendix Figure 35 Legend: Data from this figure was extracted from the study published by Nathan et al in 2006, which investigates the effect of ondansetron/metopimazine and ondansetron monotherapy on emetic episodes per day in children with brain tumors receiving highly emetogenic therapy. The average treatment effect is -0.56 (-1.74 to 0.62). White circles indicate placebo; black circles indicate metopimazine.

Appendix Figure 36: Patients with unstable angina at rest treated with oral verapamil and placebo and its effect on ischemic attacks[23]



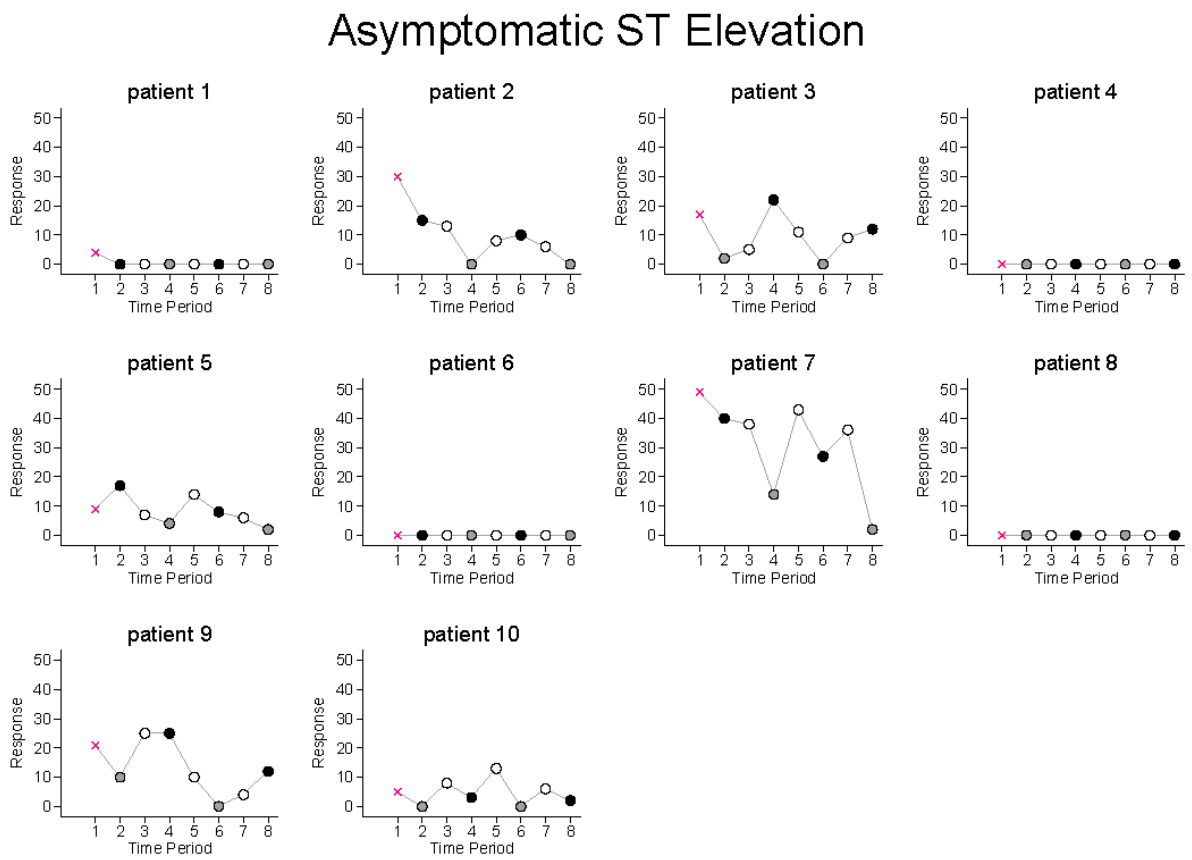
Appendix Figure 36 Legend: Data from this figure was extracted from the study published by Parodi et al in 1979, which investigates the effect of oral verapamil and placebo on ischemic attacks in patients with unstable angina. The average treatment effect is -1.63 (-2.10 to -1.17). Red circles indicate baseline; white circles indicate placebo; black circles indicate verapamil.

Appendix Figure 37: Patients with unstable angina at rest treated with verapamil, propranolol and placebo and its effect on asymptomatic ST depression[24]



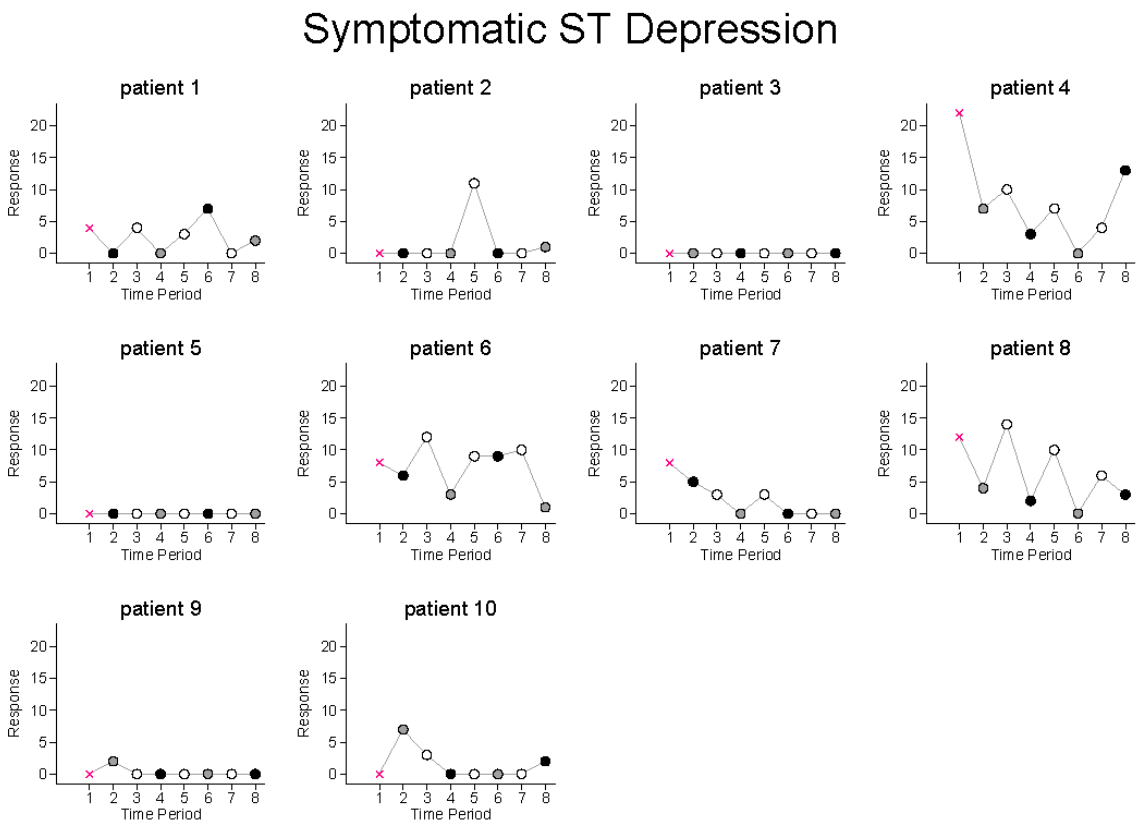
Appendix Figure 37 Legend: Data from this figure was extracted from the study published by Parodi et al in 1986, which investigates the effect of verapamil, propranolol and placebo on asymptomatic ST depression in patients with unstable angina. The average treatment effect is -0.82 (-2.54 to 0.90). Red Xs indicate baseline; white circles indicate placebo; grey circles indicate propranolol; black circles indicate verapamil.

Appendix Figure 38: Patients with unstable angina at rest treated with verapamil, propranolol and placebo and its effect on asymptomatic ST elevation[24]



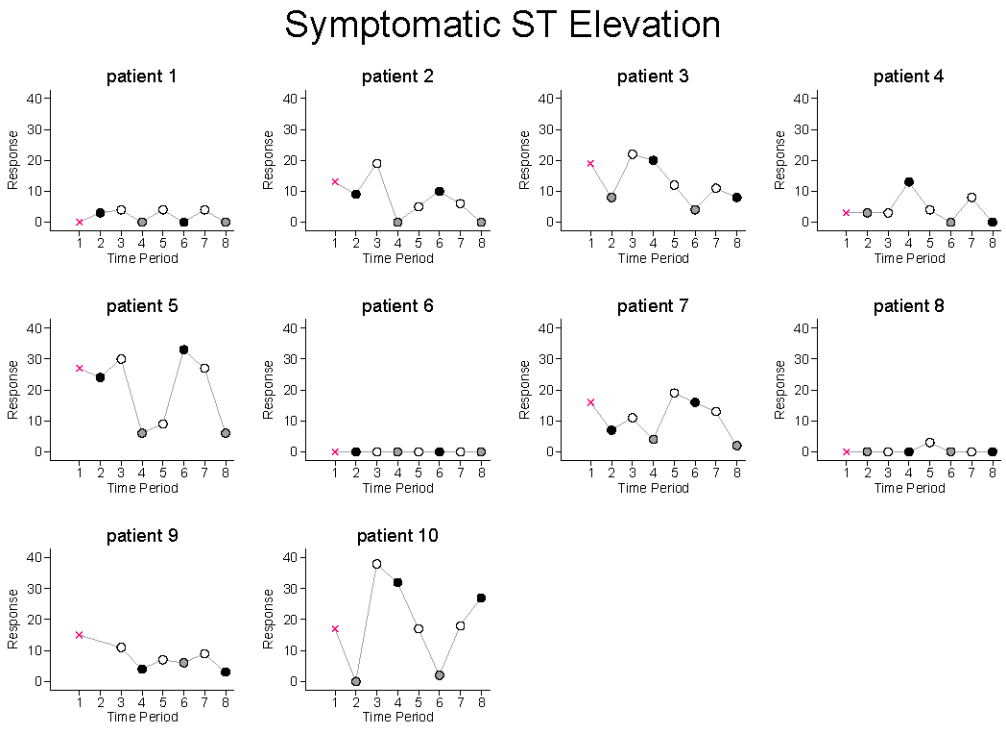
Appendix Figure 38 Legend: Data from this figure was extracted from the study published by Parodi et al in 1986, which investigates the effect of verapamil, propranolol and placebo on asymptomatic ST elevation in patients with unstable angina. The average treatment effect is -1.97 (-2.92 to -1.01). Red Xs indicate baseline; white circles indicate placebo; grey circles indicate propranolol; black circles indicate verapamil.

Appendix Figure 39: Patients with unstable angina at rest treated with verapamil, propranolol and placebo and its effect on symptomatic ST depression[24]



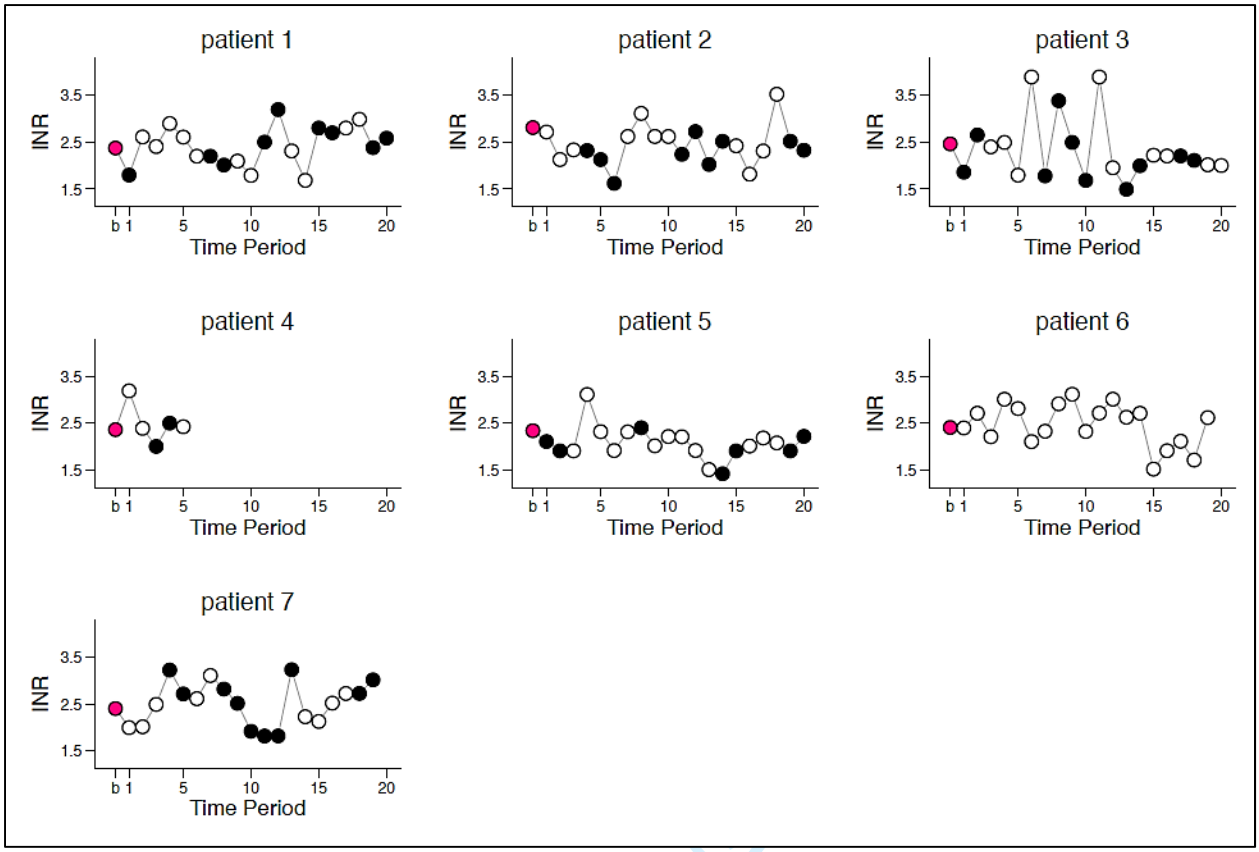
Appendix Figure 39 Legend: Data from this figure was extracted from the study published by Parodi et al in 1986, which investigates the effect of verapamil, propranolol and placebo on symptomatic ST depression in patients with unstable angina. The average treatment effect is -0.98 (-1.84 to -0.13). Red Xs indicate baseline; white circles indicate placebo; grey circles indicate propranolol; black circles indicate verapamil.

Appendix Figure 40: Patients with unstable angina at rest treated with verapamil, propranolol and placebo and its effect on symptomatic ST elevation[24]



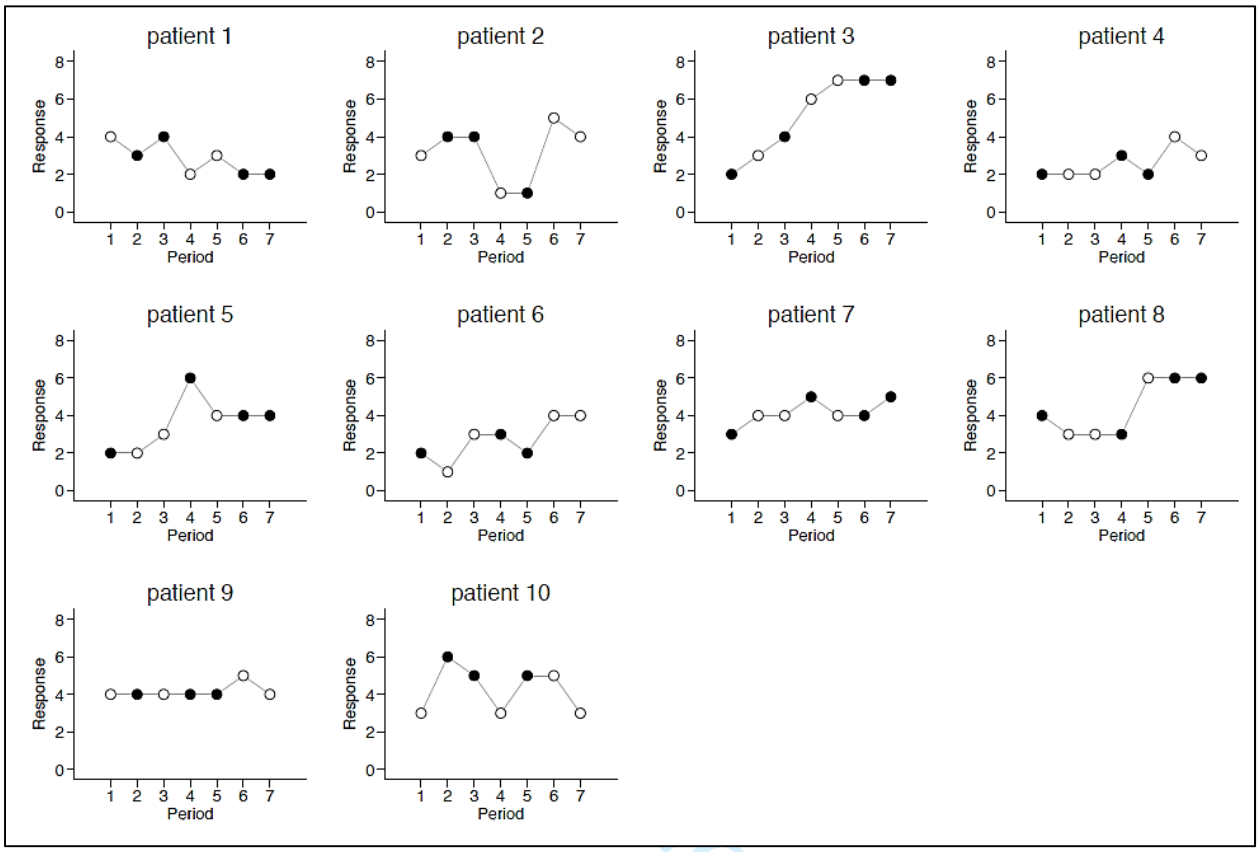
Appendix Figure 40 Legend: Data from this figure was extracted from the study published by Parodi et al in 1986, which investigates the effect of verapamil, propranolol and placebo on symptomatic ST elevation in patients with unstable angina. The average treatment effect is -1.87 (-2.72 to -1.02). Red Xs indicate baseline; white circles indicate placebo; grey circles indicate propranolol; black circles indicate verapamil.

Appendix Figure 41: Patients previously taking warfarin for either atrial fibrillation or deep vein thrombosis treated with apo-warfarin and coumadin and its effect on international normalized ratio[8]



Appendix Figure 41 Legend: Data from this figure was extracted from the study published by Pereira et al in 1995, which investigates the effect of apo-warfarin and coumadin on international normalized ratio in patients previously taking warfarin for either atrial fibrillation or deep vein thrombosis. The average treatment effect is -0.12 (-0.30 to 0.07). Red circles indicate baseline; white circles indicate Coumadin; black circles indicate apo-warfarin.

Appendix Figure 42: Parkinson's disease patients with troublesome dyskinesia treated with simvastatin and placebo and its effect on discomfort caused by troublesome dyskinesia[25]



Appendix Figure 42 Legend: Data from this figure was extracted from the study published by Tison et al in 2012, which investigates the effect of simvastatin and placebo on discomfort caused by troublesome dyskinesia in Parkinson's disease patients with troublesome dyskinesia. The average treatment effect is0.20 (-0.40 to 0.80). White circles indicate placebo; black circles indicate simvastatin.

Appendix Reference List

(1) Emmanuel AV, Kamm MA, Roy AJ, Kerstens R, Vandeplassche L. Randomised clinical trial: the efficacy of prucalopride in patients with chronic intestinal pseudo-obstruction--a double-blind, placebo-controlled, cross-over, multiple n = 1 study. *Aliment Pharmacol Ther* 2012; 35(1):48-55.

(2) Haas DC, Sheehe PR. Dextroamphetamine pilot crossover trials and n of 1 trials in patients with chronic tension-type and migraine headache. *Headache* 2004; 44(10):1029-1037.

(3) Jaeschke R, Adachi J, Guyatt G, Keller J, Wong B. Clinical usefulness of amitriptyline in fibromyalgia: the results of 23 N-of-1 randomized controlled trials. *J Rheumatol* 1991; 18(3):447-451.

(4) Johannessen T, Petersen H, Kristensen P, Fosstvedt D, Kleveland PM, Dybdahl J et al. Cimetidine on-demand in dyspepsia. Experience with randomized controlled single-subject trials. *Scand J Gastroenterol* 1992; 27(3):189-195.

(5) Mahon J, Laupacis A, Donner A, Wood T. Randomised study of n of 1 trials versus standard practice. *BMJ* 1996; 312(7038):1069-1074.

(6) March L, Irwig L, Schwarz J, Simpson J, Chock C, Brooks P. n of 1 trials comparing a non-steroidal anti-inflammatory drug with paracetamol in osteoarthritis. *BMJ* 1994; 309(6961):1041-1045.

(7) Patel A, Jaeschke R, Guyatt GH, Keller JL, Newhouse MT. Clinical usefulness of n-of-1 randomized controlled trials in patients with nonreversible chronic airflow limitation. *Am Rev Respir Dis* 1991; 144(4):962-964.

(8) Pereira JA, Holbrook AM, Dolovich L, Goldsmith C, Thabane L, Douketis JD et al. Are brand-name and generic warfarin interchangeable? Multiple n-of-1 randomized, crossover trials. *Ann Pharmacother* 2005; 39(7-8):1188-1193.

(9) Wallace AE, Kofoed LL. Statistical analysis of single case studies in the clinical setting: the example of methylphenidate trials in children with attention-deficit hyperactivity disorder. *J Child Adolesc Psychopharmacol* 1994; 4(3):141-150.

(10) Woodfield R, Goodyear-Smith F, Arroll B. N-of-1 trials of quinine efficacy in skeletal muscle cramps of the leg. *Br J Gen Pract* 2005; 55(512):181-185.

(11) Zucker DR, Ruthazer R, Schmid CH, Feuer JM, Fischer PA, Kieval RI et al. Lessons learned combining N-of-1 trials to assess fibromyalgia therapies. *J Rheumatol* 2006; 33(10):2069-2077.

(12) Joy TR, Monjed A, Zou GY, Hegele RA, McDonald CG, Mahon JL. N-of-1 (single-patient) trials for statin-related myalgia. *Ann Intern Med* 2014; 160(5):301-310.

(13) Lipka AF, Vrinten C, van Zwet EW, Schimmel KJ, Cornel MC, Kuijpers MR et al. Ephedrine treatment for autoimmune myasthenia gravis. *Neuromuscul Disord* 2017; 27(3):259-265.

(14) Camfield P, Gordon K, Dooley J, Camfield C. Melatonin appears ineffective in children with intellectual deficits and fragmented sleep: six "N of 1" trials. *J Child Neurol* 1996; 11(4):341-343.

(15) Hinderer SR. The supraspinal anxiolytic effect of baclofen for spasticity reduction. *Am J Phys Med Rehabil* 1990; 69(5):254-258.

(16) Langer JC, Winthrop AL, Issenman RM. The single-subject randomized trial. A useful clinical tool for assessing therapeutic efficacy in pediatric practice. *Clin Pediatr (Phila)* 1993; 32(11):654-657.

(17) Lashner BA, Hanauer SB, Silverstein MD. Testing nicotine gum for ulcerative colitis patients. Experience with single-patient trials. *Dig Dis Sci* 1990; 35(7):827-832.

(18) Maier W, Benkert O. Treatment of chronic depression with sulpiride: evidence of efficacy in placebo-controlled single case studies. *Psychopharmacology (Berl)* 1994; 115(4):495-501.

(19) Mandelcorn J, Cullen NK, Bayley MT. A preliminary study of the efficacy of ondansetron in the treatment of ataxia, poor balance and incoordination from brain injury. *Brain Inj* 2004; 18(10):1025-1039.

(20) McQuay HJ, Carroll D, Jadad AR, Glynn CJ, Jack T, Moore RA et al. Dextromethorphan for the treatment of neuropathic pain: a double-blind randomised controlled crossover trial with integral n-of-1 design. *Pain* 1994; 59(1):127-133.

(21) Miyazaki S, Nonogi H, Goto Y, Sumiyoshi T, Haze K, Hiramori K. Comparison of the therapeutic efficacy of continuous and intermittent injection of isosorbide dinitrate: a randomized study on unstable angina. *Intern Med* 1995; 34(9):856-862.

(22) Nathan PC, Tomlinson G, Dupuis LL, Greenberg ML, Ota S, Bartels U et al. A pilot study of ondansetron plus metopimazine vs. ondansetron monotherapy in children receiving highly emetogenic chemotherapy: a Bayesian randomized serial N-of-1 trials design. *Support Care Cancer* 2006; 14(3):268-276.

(23) Parodi O, Maseri A, Simonetti I. Management of unstable angina at rest by verapamil. A double-blind cross-over study in coronary care unit. *Br Heart J* 1979; 41(2):167-174.

(24) Parodi O, Simonetti I, Michelassi C, Carpeggiani C, Biagini A, L'Abbate A et al. Comparison of verapamil and propranolol therapy for angina pectoris at rest: a randomized, multiple-crossover, controlled trial in the coronary care unit. *Am J Cardiol* 1986; 57(11):899-906.

(25) Tison F, Negre-Pages L, Meissner WG, Dupouy S, Li Q, Thiolat ML et al. Simvastatin decreases levodopa-induced dyskinesia in monkeys, but not in a randomized, placebo-controlled, multiple cross-over ("n-of-1") exploratory trial of simvastatin against levodopa-induced dyskinesia in Parkinson's disease patients. *Parkinsonism Relat Disord* 2013; 19(4):416-421.

(26) Joy TR, Monjed A, Zou GY, Hegele RA, McDonald CG, Mahon JL. N-of-1 (single-patient) trials for statin-related myalgia. *Ann Intern Med* 2014; 160(5):301-310.

Appendix Table 5. Studies reporting person-level treatment effect with both fixed-effect and random-effect using a method of moments estimator

Study	Outcome	Fixed effect model	P for HTE (fixed-effects model)	Random Treatment Effect	summary_tau2	P for HTE (random-effects model)
March 1994[6]	Mean pain score on VAS taken from 2nd week of tx	-4.155 (-4.807 to -3.502)	<0.001	-7.093 (-11.939 to -2.248)	73.530	<0.001
March 1994[6]	Mean stiffness score on VAS taken from 2nd week of	-2.192 (-2.549 to -1.835)	<0.001	-5.992 (-11.280 to -0.704)	88.872	<0.001
Emmanuel 2012[1]	Bloating	-0.131 (-0.171 to -0.090)	<0.001	-0.344 (-0.619 to -0.069)	0.071	<0.001
Emmanuel 2012[1]	Pain	-0.160 (-0.209 to -0.111)	<0.001	-0.440 (-0.771 to -0.110)	0.106	<0.001
Haas 2004[2]	Chronic tension-type headache grade	0.733 (0.609 to 0.857)	<0.001	0.772 (0.454 to 1.090)	0.350	<0.001
Haas 2004[2]	Chronic tension-type headache grade	0.543 (0.394 to 0.693)	0.067	0.542 (0.354 to 0.731)	0.055	0.067
Jaeschke 1991[3]	7-point symptom scale	0.356 (0.286 to 0.426)	<0.001	0.427 (0.210 to 0.645)	0.186	<0.001
Jaeschke 1991[3]	Tender point changes count	1.072 (0.701 to 1.443)	<0.001	1.320 (0.404 to 2.236)	2.166	<0.001
Johannessen 1992[4]	6-point symptom scale	0.657 (0.530 to 0.785)	<0.001	0.698 (0.466 to 0.931)	0.382	<0.001
Joy 2014[26]	VAS myalgia score	0.119 (-2.283 to 2.521)	0.995	0.119 (-2.283 to 2.521)	0.000	0.996
Joy 2014[26]	Symptom-specific VAS	1.937 (0.179 to 3.696)	0.797	1.937 (0.179 to 3.696)	0.000	0.797
Joy 2014[26]	Pain severity score	0.086 (-0.215 to 0.387)	0.986	0.086 (-0.215 to 0.387)	0.000	0.986

Joy 2014[26]	Pain interference score	-0.016 (-0.095 to 0.064)	0.917	-0.016 (-0.095 to 0.064)	0.000	0.917
Lipka 2017[13]	Quantitative myasthenia gravis score	1.006 (0.215 to 1.797)	0.803	1.006 (0.215 to 1.797)	0.000	0.803
Lipka 2017[13]	Myasthenia gravis composite	2.952 (0.969 to 4.934)	0.177	2.891 (0.348 to 5.433)	2.631	0.177
Lipka 2017[13]	MG-ADL	1.110 (0.269 to 1.951)	0.047	1.099 (-0.277 to 2.474)	1.222	0.047
Lipka 2017[13]	VAS score	1.204 (0.124 to 2.283)	0.190	1.275 (-0.115 to 2.665)	0.739	0.190
Mahon 1996[5]	Likert Scale (1-7)	0.069 (-0.042 to 0.179)	<0.001	0.145 (-0.153 to 0.443)	0.134	<0.001
Patel 1991[7]	4-item symptom questionnaire	0.000 (-0.000 to 0.000)	<0.001	0.000 (-0.000 to 0.000)	0.000	<0.001
Pereira 1995[8]	INR (diff)	0.027 (-0.155 to 0.209)	0.477	0.027 (-0.155 to 0.209)	0.000	0.477
Wallace 1994[9]	Conners 15-item rating scale scores	0.759 (0.341 to 1.178)	0.747	0.759 (0.341 to 1.178)	0.000	0.747
Woodfield 2005[10]	Number of cramps	-5.395 (-7.091 to -3.699)	<0.001	-18.823 (-28.527 to -9.120)	161.582	<0.001
Woodfield 2005[10]	Total days with cramps	-7.600 (-8.420 to -6.781)	<0.001	-6.181 (-9.798 to -2.563)	26.245	<0.001
Zucker 2006[11]	FIQ	-5.019 (-8.784 to -1.254)	0.999	-5.019 (-8.784 to -1.254)	0.000	0.999

Appendix Table 6. Studies reporting person-level outcomes with both fixed-effect and random-effect hierarchical linear model

Author Year	Outcome	Range of the Scales (severity)	Fixed Treatment Effect	Random Treatment Effect	P-value Person Treatment Interaction
Camfield 1996[14]	Nights without awakening	NR	0.865 (0.215 to 1.516)	0.84 (0.20 to 1.48)	0.456
Hinderer 1990[15]	Anxiety	Beck Inventory-A anxiety scale 0-3 (0 = never, 3 = almost all the time)	0.000 (0.000 to 0.000)	-1.06 (-1.88 to -0.23)	<0.001
Joy 2014[26]	Myalgia score	Visual Analogue Score for myalgia (0=none to 100=worst)	3.3812 (-2.668 to 9.430)	3.3522 (-2.617 to 9.322)	0.566
Langer 1993[16]	Vomiting	NR	-1.204 (-2.494 to 0.086)	-1.20 (-2.49 to 0.09)	0.136
Lashner 1990[17]	Symptom score: abdominal pain	Symptom scores 0-100 (0=best, 100=worst)	-3.615 (-16.982 to 9.751)	-3.62 (-15.84 to 8.61)	0.007
	Symptom score: bowel movements/day		-0.538 (-1.215 to 0.138)	-0.56 (-1.22 to 0.09)	0.001
	Symptom score: consistency of bowel movements		7.000 (-7.551 to 21.551)	7.00 (-6.29 to 20.29)	0.013
	Symptom score: hematochezia		2.308 (-17.210 to 21.826)	2.35 (-17.21 to 21.90)	0.003
	Symptom score: general sense of well-being		-6.538 (-25.352 to 12.275)	-6.54 (-23.62 to 10.56)	0.008
Maier 1994[18]	SCL-90 subscales: Depressed mood	NR	-3.536 (-6.718 to -0.354)	-3.63 (-7.40 to 0.15)	<0.001
	SCL-90 subscales: Anxiety		-3.753 (-6.582 to -0.924)	-3.81 (-7.22 to -0.40)	<0.001
	SCL-90 subscales: Somatization		-1.419 (-4.316 to 1.478)	-1.50 (-4.20 to 1.21)	0.869
Mandelcorn 2004[19]	Self-Assessment score	0–5 (0=worst, 5=best)	-2.052 (-8.865 to 4.761)	-2.05 (-8.43 to 4.33)	0.05
	Lower extremity ataxia	Fugl-Meyer: 3-point (0 cannot be performed to 2 can	12.494 (-3.155 to 28.142)	12.49 (-0.85 to 25.84)	0.025

Author Year	Outcome	Range of the Scales (severity)	Fixed Treatment Effect	Random Treatment Effect	P-value Person Treatment Interaction
		be fully performed)			
	Truncal ataxia	AMTI forceplate®: NR *Berg Balance Scale® 0–56, with a higher score indicating a better performance*	1.196 (-2.866 to 5.257)	1.20 (-2.06 to 4.45)	0.690
	Upper extremity ataxia	Purdue Pegboard Test®: pegs inserted into the board with each hand in 30 sec *Minnesota Placing Test®: reach out, grasp, and place blocks in a specific order*	-0.498 (-3.546 to 2.550)	-0.50 (-3.10 to 2.10)	0.382
McQuay 1994[20]	VAS Pain Intensity	0-100 (0 = no pain, 100 = worst possible pain)	-1.094 (-5.572 to 3.383)	-1.06 (-5.16 to 3.04)	0.004
	VAS Relief Intensity	0-100 (0 = no relief, 100 =complete pain relief)	-3.913 (-11.729 to 3.903)	-3.86 (-11.11 to 3.40)	0.038
Miyazaki 1995[21]	Incidence of angina	Either ST-segment elevation or depression at rest	0.496 (-0.206 to 1.199)	0.47 (-0.32 to 1.26)	0.125
Nathan 2006[22]	Emetic episodes per day	complete response (0 episodes/day), major response (1–2 episodes/day), or failure (>2 episodes/day)	-0.095 (-0.514 to 0.325)	-0.56 (-1.74 to 0.62)	0.001
Parodi 1979[23]	Ischemic attacks	ST elevation or depression (details NR)	-1.544 (-1.838 to -1.251)	-1.63 (-2.10 to -1.17)	0.007
Parodi 1986[24]	Asymptomatic ST elevation (After verapamil)	NR	-1.637 (-1.994 to -1.279)	-1.97 (-2.92 to -1.01)	0.110
	Asymptomatic ST depression (After verapamil)		-1.083 (-1.903 to -0.262)	-0.82 (-2.54 to 0.90)	0.401

Author Year	Outcome	Range of the Scales (severity)	Fixed Treatment Effect	Random Treatment Effect	P-value Person Treatment Interaction
	Symptomatic ST elevation (After verapamil)		-1.580 (-1.906 to -1.254)	-1.87 (-2.72 to -1.02)	<0.001
	Symptomatic ST Depression (After verapamil)		-0.990 (-1.411 to -0.569)	-0.98 (-1.84 to -0.13)	0.002
	Asymptomatic ST elevation (After propranolol)		0.100 (-0.086 to 0.286)	-1.966 (-2.917 to -1.014)	0.006
	Asymptomatic ST depression (After propranolol)		0.339 (-0.168 to 0.845)	-0.821 (-2.539 to 0.897)	0.964
	Symptomatic ST elevation (After propranolol)		-0.002 (-0.177 to 0.173)	-1.868 (-2.718 to -1.017)	0.063
	Symptomatic ST Depression (After propranolol)		-0.374 (-0.709 to -0.039)	-0.981 (-1.835 to -0.126)	0.023
Pereira 1995[8]	INR	Target INR range of 2.0–3.0		-0.12 (-0.30 to 0.07)	0.433
Tison 2012[25]	Troublesome dyskinesia	7 points scale (1=extremely uncomfortable, 7=not at all uncomfortable)		0.20 (-0.40 to 0.80)	0.593

Statistical codes for analysis results of studies reporting person-level treatment effects

Estimation of standard errors in the following studies

- Emmanuel 2012: gen SE_Intervention (or control) =SD of intervention (or control) score/square root of Intervention days (or control days)
- Haas 2004: SE was available in Table 4 of the original paper
- Jaeschke 1991, Patel 1991, March 1994, Woodfield 2005, Wallace 1994 - SE was derived using the p-value of one-sided paired t-test of the difference in score using the following code:
  generate t_stat = invt(2,p_value)
  generate se = abs(mean_outcome/t_stat)
- Johannessen 1992, Pereira 1995, Zucker 2006, Joy 2014, Lipka 2017 – SE was derived from the 95% confidence interval using the following code: generate se = (UCI - LCI) /(2*invnorm(0.975))
- Mahon 1996: SE was derived from 95% confidence interval based on Student's t distribution using the following code: generate se = (UCI - LCI) /(2*invt(DF, 0.975))

```
metan difference se_difference  if Outcome == "outcome", random        **/fixedi is used for fixed effect model

    local p = r(p_het)

    local sum_es =  r(ES)

    local sum_es_se = r(seES)

    local tau2=  r(tau2)

    local I_sq =  r(i_sq)

post `memory' ("`study'") ("`outcome'")  (`sum_es') (`sum_es_se')  (`tau2') (`I_sq') (`p')
```

Statistical codes for analysis results of studies reporting person-level outcome effects

```
egen id = group(Patient)
generate tx = 0 if Exposure == "Placebo"
        replace tx = 1 if Exposure == "Intervention"
egen period_seq = seq(), from(1) to(18) */varies based on the number of periods*/
local outcome = "Specific_outcome"
        /* fixed baselines and random treatment effects */
        xtmixed Result  tx i.id || id: tx  if Outcome == "`outcome'"  , nocons
                estimates store D
                matrix estimates = e(b)
                local point_estimate_ran_bas_ran_tx = estimates[1,1]
                local sd_estimate_rand_base_random_tx = (exp(estimates[1,10]))

                matrix variances = e(V)
                local point_se_rand_base_random_tx = sqrt(variances[1,1])
                local point_low_ran_bas_ran_tx = `point_estimate_ran_bas_ran_tx' - invnormal(0.975) * `point_se_rand_base_random_tx'
                local point_up_ran_bas_ran_tx = `point_estimate_ran_bas_ran_tx' + invnormal(0.975) * `point_se_rand_base_random_tx'

                local sd_se_rand_base_random_tx = sqrt(variances[10,10])
```

```
		local sd_lower_rand_base_random_tx = (exp(ln((`sd_estimate_rand_base_random_tx')) - invnormal(0.975) *
`sd_se_rand_base_random_tx'))

		local sd_upper_rand_base_random_tx = (exp(ln((`sd_estimate_rand_base_random_tx')) + invnormal(0.975) *
`sd_se_rand_base_random_tx'))


	/* fixed baselines and common treatment effect -- linear regression */

	xtmixed Result  tx i.id  || id: if Outcome == "`outcome'" , nocons

		estimates store E


	/* fixed baselines and person interactions */

	regress Result i.tx##i.id if Outcome == "`outcome'"

		estimates store F


	/* fixed baselines and common effects */

	regress Result tx i.id if Outcome == "`outcome'"

		estimates store G


		matrix estimates = e(b)

		local point_estimate_fix_bas_com_tx = estimates[1,1]
```

```
            matrix variances = e(V)
            local point_se_fix_bas_common_tx = sqrt(variances[1,1])
            local t_stat = `point_estimate_fix_bas_com_tx' /  `point_se_fix_bas_common_tx'
            local point_low_fix_bas_com_tx = `point_estimate_fix_bas_com_tx' - invt(e(df_r), 0.975) * `point_se_fix_bas_common_tx'
            local point_up_fix_bas_com_tx = `point_estimate_fix_bas_com_tx' + invt(e(df_r), 0.975) * `point_se_fix_bas_common_tx'

        lrtest D E
            local p_random_RANDOM_FIXED_tx =  r(p)

        lrtest F G
            local p_person_by_treat =  r(p)

        post `memory'   ("Study") ("`outcome'")
```

Please note: Depending on the outcome, xtmixed or meqrlogit or meqrpoisson was used.

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	1-2
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	2
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	n/a
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5-6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5-6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	a1-a3
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	6-7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	6-8
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	n/a
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	8-9
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	8-9

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	n/a
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	8-9
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	9, 20, 21, 29
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	1-12, 22-26
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	n/a
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	31, a11-a50
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	10-12, 26
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	n/a
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	12, a53-a57
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	12-14
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	15-16
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	16
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	17

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

PRISMA 2009 Checklist

BMJ Open

Evaluation of person-level heterogeneity of treatment effects in published multi-person N-of-1 studies: systematic review and re-analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-017641.R2
Article Type:	Research
Date Submitted by the Author:	22-Feb-2018
Complete List of Authors:	Raman, G; Tufts Medical Center Balk, EM; Brown University Lai, Lana; Tufts Medical Center Shi, Jennifer; Tufts Medical Center Chan, Jeffrey; VA Boston Healthcare System, Center for Healthcare Organization and Implementation Research (CHOIR) Lutz, Jennifer; Tufts Medical Center Dubois, Robert; National Pharmaceutical Council, Research Kravitz, Richard; University of California Davis Kent, David; Tufts Medical Center
Primary Subject Heading:	Patient-centred medicine
Secondary Subject Heading:	Research methods
Keywords:	perseonalized medicine, n-of-1 studies, systematic review, heterogeneity of treatment effect

Evaluation of person-level heterogeneity of treatment effects in published multi-person N-of-1 studies: systematic review and re-analysis

Gowri Raman, MD, MS[a]; Ethan M Balk, MD, MPH[b]; Lana Lai, MS[c]; Jennifer Shi, BA[d]; Jeff Chan, MD[a,e]; Jennifer Lutz, MA[c]; Robert Dubois, MD, PhD[f]; Richard L Kravitz, MD, MSPH[g]; David M Kent, MD, MS*[c]

[a]Center for Clinical Evidence Synthesis, ICRHPS, Medical Center/Tufts University School of Medicine, Boston, USA; [b]Center for Evidence Synthesis in Health, School of Public Health, Brown University, Providence, RI; [c]Predictive Analytics and Comparative Effectiveness (PACE) Center, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center/Tufts University School of Medicine, Boston, USA; [d]Center for the Evaluation of Value and Risk in Health, ICRHPS, Medical Center/Tufts University School of Medicine, Boston, USA; [e] VA Boston Healthcare System, Center for Healthcare Organization and Implementation Research (CHOIR); [f]National Pharmaceutical Council; [g]Department of Internal Medicine, University of California, Davis

Corresponding Author:
David Kent, MD, MS
Director, Predictive Analytics and Comparative Effectiveness (PACE) Center
Institute for Clinical Research and Health Policy Studies
Tufts Medical Center
800 Washington St, Box 63
Boston, MA 02111
Email: dkent1@tuftsmedicalcenter.org
Phone: 617-636-3234

Running title: Variation in person-level treatment effects: systematic review
Word count
Abstract: 224
Main text: 4,259 (main text, references)
Table: 5
Figures: 3
Key words: n-of-1 studies, systematic review, heterogeneity of treatment effect, personalized medicine

Abstract

Objective: Individual patients with the same condition may respond differently to similar treatments. Our aim is to summarize the reporting of person-level heterogeneity of treatment effects (HTE) in multi-person N-of-1 studies and to examine the evidence for person-level HTE through re-analysis.

Study Design: Systematic review and re-analysis of multi-person N-of-1 studies.

Data sources: Medline, Cochrane Controlled Trials, EMBASE, Web of Science, and review of references through August 2017 for N-of-1 studies published in English.

Study Selection: N-of-1 studies of pharmacological interventions with at least two subjects.

Data Synthesis: Citation screening and data extractions were performed in duplicate. We performed statistical reanalysis testing for person-level HTE on all studies presenting person-level data.

Results: We identified 62 multi-person N-of-1 studies with at least two subjects. Statistical tests examining HTE were described in only 13 (21%), of which only two (3%) tested person-level HTE. Only 25 studies (40%) provided person-level data sufficient to re-analyze person-level HTE. Reanalysis using a fixed effect linear model identified statistically significant person-level HTE in 8 of the 13 studies (62%) reporting person-level treatment effects and in 8 of the 14 studies (57%) reporting person-level outcomes.

Conclusions: Our analysis suggests person-level HTE is common and often substantial. Reviewed studies had incomplete information on person-level treatment effects and their variation. Improved assessment and reporting of person-level treatment effects in multi-person N-of-1 studies are needed.

Strengths and limitations of this study

- Multi-person N-of-1 studies are the best design to estimate individual patient treatment effects and compare the variation in effects between individuals to variation within individuals across different periods.
- Our analysis suggests person-level HTE is common and often substantial.
- Our analysis was limited by the paucity of N-of-1 studies in the literature and by the low statistical power in the available studies.

Introduction

Clinicians commonly observe that individual patients given the same treatment for the same condition appear to respond differently from one another. This observation, combined with our understanding of the complex mechanisms of diseases and therapies and the potential importance of myriad patient-specific factors (e.g., age, sex, illness severity, comorbidities, co-treatments, and molecular differences influencing pharmacokinetics and -dynamics), have led to a widely held assumption that the observed variation in treatment response seen between individuals is not merely random, but stable and potentially predictable. This assumption underpins the field of personalized medicine, which aims to determine the best treatment for an individual patient, as opposed to treating all patients with the intervention found to be most effective for the "average" patient.

Nevertheless, statistical analyses aimed at discovering heterogeneity of treatment effects (HTE) among groups of individuals (for example subgroup analyses of parallel arm randomized trials) typically fail to find compelling and reliable evidence for the presence of such heterogeneity. For example, statistically significant differences in treatment effects between men and women are often reported, but a systematic review indicates that the frequency of these interactions across studies suggests the vast majority occur by chance.[1] Similarly, the field of pharmacogenetics, also built on the assumption of stable variation in treatment responses, has largely failed to live up to its promise to broadly improve the targeting of drugs—particularly outside the special case of oncology (where studies generally depend on the subclassification of tumor tissue not on variation in germline polymorphisms).[2;3] This failure to find reproducible HTE has supported the contrarian notion that true individual effects may be a "myth," an over-interpretation of random noise.[4]

To distinguish between these two possibilities, Kalow et al. have suggested that carefully designed series of N-of-1 studies could be performed for those chronic conditions amenable to this design (i.e., where the disease process is relatively stable over time, treatment effects are transient, and outcomes vary and are observable over time).[5] By estimating individual patient treatment effects and comparing the variation in effects *between* individuals to variation *within* individuals across different periods, it is possible to determine the non-random component of heterogeneity in individual treatment effects--even if one is unable to identify the variables that predict this variation (i.e., even in the absence of group-level HTE, such as men versus women, or old versus young).

A recent review summarized N-of-1 studies reported in the literature—including multi-person N-of-1 studies—but did not examine whether and how these studies provide information on person-level HTE. Therefore our objectives are: 1) to summarize the conduct and reporting of assessments of variation in person-level treatment effects from N-of-1 studies; and 2) to extract, reanalyze and report the results from the subset of studies that provided adequate data in their published reports to examine the extent of the evidence for person-level HTE (i.e., participant-level outcomes or effects).[6]

Methods

This review was conducted in accordance with the highest standards for conducing systematic reviews.[7;8] We defined N-of-1 studies as crossover trials in which each patient receives two or more treatments in a pre-defined, often randomized, sequence.

Data Sources and Searches

We used two separate searches because N-of-1 studies can be indexed differently: (1) a search in Medline, Cochrane Central and EMBASE using terms related to repeated crossover studies (for publications indexed from inception to August 17, 2017); and (2) a Medline, Cochrane Central, EMBASE, and Web of Science search using terms that are related to N-of-1 (for publications indexed from 2011 to August 17, 2017). For N-of-1 studies indexed before 2011, we used studies included in a prior published systematic review by Gabler et al.[6] Our searches combined terms and Medical Subject Headings for N-of-1, single-subject, single-patient, randomized trials, crossover, multi-period crossover, and rotated or repeated period crossover (see Appendix Tables 1-2 for detailed search terms). The searches were not restricted by disease, condition, organ system, or treatment.

Study Selection

We selected eligible multi-person N-of-1 studies to describe the frequency of reporting of individual outcomes and effects and of documented HTE in these studies. We required a minimum of two individual subjects per study for evaluation of HTE. We excluded studies that included non-pharmacological interventions, reviews, abstracts and protocols. We include studies with placebo or "no treatment" interventions. Citations were double-screened by reviewers using an open-source, online software Abstrackr (http://abstrackr.cebm.brown.edu/). Full-text articles of potentially relevant studies were again double screened for eligibility.

Person-level outcomes were defined as outcomes for each person at each point in time when they were measured, reported in tables, text, or graphs. Person-level treatment effect was defined as contrasts of outcomes in individuals on one treatment versus the comparator. Person-level HTE was defined as quantified variation in the person-level treatment effects, whereas HTE

more broadly includes any type of subgroup analysis (e.g., males versus females; older versus younger) as outlined in **Figure 1**.

Data Extraction and Quality Assessment

One of four reviewers extracted data from each publication; a second reviewer verified all numerical information and basic descriptors of the study design and analysis. Operational definitions for extraction items were discussed in weekly project meetings and discrepancies between extractors were resolved by consensus with senior authors (DK, GR, EB). From each study, we extracted bibliographic information, details related to study design (number of patients enrolled, selection criteria, interventions evaluated, randomization methods, outcomes assessed, follow-up duration), information on patient characteristics, and person-level measurements of outcomes or estimates of person-level treatment effects (with corresponding measures of their uncertainty). When necessary, we extracted data by digitizing the graphs and the values were estimated using Engauge Digitizer version 2.14 (http://digitizer.sourceforge.net/). We assessed the methodological quality of each study based on predefined criteria, in accordance with the Agency for Healthcare Research and Quality (AHRQ) suggested methods and the Cochrane risk of bias for clinical trials.[9;10]

We generated graphs showing the trajectory of response for each patient in each study and compared them against the published information. We also generated scatterplots of measurements over time for studies that did not present their data in graphical format to help us identify aberrant data points (e.g., errors in data extraction). We verified potentially aberrant data points by re-examining the published data and made corrections, when needed.

Data Synthesis and Analyses

We examined the degree to which studies reported person-level data. This was described using the following items for each reported outcome: 1) qualitative descriptions of HTE (e.g., "there were 8 responders and 4 non-responders"); 2) details of person-level outcomes (i.e., outcomes with each treatment within each period); 3) details of person-level treatment effect (i.e., a point estimate of contrasts of outcomes in individuals on one treatment versus the comparator); 4) reporting of person-level statistical effect estimate, (e.g., standard deviation, exact P values, or confidence intervals for treatment effects within individuals); 5) description of statistical tests examining HTE (i.e., tests evaluating the contrast of treatment effects between individuals or groups in the study); and 6) claims of HTE. Note that qualitative descriptions of HTE for item 1 would include any description that implied that treatment effects varied, whereas item 6 required a more definite study conclusion (e.g., "our results demonstrate significant variation across individuals in response to treatment X"), whether or not these conclusions were based on robust statistical tests.

Statistical HTE analysis of extracted study results

We performed statistical analysis testing for person-level HTE on all studies presenting person-level data. We used a consistent analytic strategy across studies, to the extent permitted by the reporting in published papers. Our strategy was different for studies that reported person-level outcome measurements and those that reported estimates of person-level treatment effects with their sampling variances (or adequate information to approximately calculate these statistics).

For studies that only reported (or allowed the calculation of) *estimates of person-level treatment effects*, we obtained an average effect using a fixed effect inverse variance model and estimated the variance of the person-level treatment effects using DerSimonian and Laird method

of moments estimator.[11;12] In addition to a fixed effect model, we also obtained an average effect using a random effects model. Finally, we tested the hypothesis that all person-level treatment effects were equal using Cochran's chi-square test and quantified the proportion of observed variation due to "true" person-level effect heterogeneity with the I^2 statistic.[13]

For studies that reported *person-level outcomes*, we developed a linear model (for continuous outcomes) or generalized linear model (for binary or count outcomes) using the outcome of interest as the response, the intervention(s) as a covariate; and indicator variables for different study participants.[4] This model estimates a common treatment effect across participants. We also derived a similar model with treatment-by-participant interactions. This model allows each patient to have a different effect. The statistical significance of person-level HTE was assessed by a likelihood ratio test comparing the two models. In addition to a fixed effect model, we also fit a hierarchical linear or generalized linear mixed model with a random intercept and a random slope (for the treatment effect) to estimate the average treatment effect across all patients (assuming person-level HTE). We tested the hypothesis that all person-level treatment effects were equal and quantified the proportion of observed variation due to 'true' person-level effect heterogeneity with the I^2 statistic.[13] For modeling within-patient variance, we used a common variance with an uncorrelated covariance structure, as was used in a prior n-of-1 study.[14] Person-level treatment effect was assumed to be equal across time-periods. For the treatment effect, we used more than one random slope when >2 treatments were compared.

Patient and Public Involvement

Patients and the public were not involved in the design or analysis of this study.

Results

The searches for repeated crossover studies identified 11,891 citations and those for N-of-1 studies identified 3819 citations (indexed from 2011 onwards). Of these, we retrieved 407 full-text articles for review plus 100 N-of-1 trial articles (indexed before 2011) from an existing systematic review.[5] Upon full-text screening, 62 studies (58 multi-person N-of-1 studies and four repeated period crossover studies) met eligibility criteria (Appendix Table 3) and are reported multi-person N-of-1 studies throughout the article. An outline of the search and study selection flow is provided in **Figure 2**.

Description of studies

Table 1 summarizes the 62 multi-person N-of-1 studies that were published between 1986 and 2017 reporting a total of 1974 patients. The most common clinical domains in the multi-person N-of-1 studies were neurology (16%), arthritis/rheumatology (10%) and psychiatry (9%). Most studies were described as "double-blind" but details about the methods for blinding were often unclear; similarly studies often provided unclear information about the generation of the randomization sequence and allocation concealment (Appendix Table 4). Among the studies, 93% compared a pair of treatment strategies, 5% compared three strategies, and 2% compared four strategies. Studies had between 3 and 16 treatment periods and obtained an average of 1 to 42 outcome measurements per period. Across reported outcomes, 89% of the assessed outcomes were patient-reported and 11% were investigator-assessed.

Reporting Person-level outcomes, effects and HTE

While most studies (92%) had some qualitative acknowledgement that the treatment effects appeared to vary across individuals, formal reporting at the participant level was variable **(Table 2)**. Person-level outcomes under each treatment were reported in 52% of multi-person N-of-1 studies. Person-level treatment effects with quantitative data (comparing outcomes on each treatment) for each individual who completed the trial was available in 32%; and details on the statistical evaluation of these effects (as standard deviations or exact P values or confidence intervals) were available in 13 (21%) multi-person N-of-1 studies. Only five (8%) studies described statistical tests examining any HTE. However, only two studies (3%) reported person-level HTE, whereas the others examined group-level HTE using conventional subgroup analysis based on observable characteristics.

Reanalysis of person-level data:

Of the 62 studies, there were 36 studies that provided person-level data, either as outcomes in each treatment period or as person-level treatment effects **(Table 3)**. Of these, only 25 studies provided person-level data sufficient to support re-analysis: 14 studies provided person-level outcomes; 13 studies provided person-level treatment effects (two studies provided both). The remaining 11 studies reported either medians or means without data on variance or did not provide sufficient information on completers, so they could not be re-analyzed for treatment effect or HTE.

Of 13 studies (with 27 unique comparisons) that reported analyzable person-level treatment effect data **(Table 3)**, 10 studies had a placebo comparator and three studies had an active comparator. The sample size ranged from 7 to 68; average crossover periods ranged from

6 to 16 days; and average outcome measures per period ranged from 1 to 21. The average treatment duration ranged from 14 to 336 days.

There were 14 studies (with 27 unique comparisons) that reported analyzable person-level outcome data **(Table 3)**, including two studies also reporting person-level treatment effects. Of these, 11 compared the intervention with placebo and three studies compared two active interventions. The sample size ranged from 2 to 22; the average number of crossover periods ranged from 3 to 10; and the average number of outcome measures per period ranged from 1 to 42. The average treatment duration ranged from 9 to 210 days.

Re-analysis of studies reporting estimates of person-level treatment effects

Thirteen studies (including 27 comparisons, due to multiple outcomes in some studies) reported estimates of person-level treatment effects sufficient to analyze (Appendix Figures 1-16 displays graphs of the person level treatment effect data). Average fixed effect estimates for each analysis are shown in **Table 4**; random effects estimates were generally similar (Appendix table 5). In 8 of the 13 studies (62%) and 15 of the 27 total unique comparisons (56%) we found evidence of statistically significant HTE for at least one outcome (Table 4). Generally, the magnitude in the variation of individual patient effects (as seen in the range) was very large compared to the average effects. Most studies (64%) showed person-level effects that differed qualitatively from one another. Most of the variation in the observed individual effects was attributable to "true" (non-random) heterogeneity of person-level effects; 11 of 27 analyses had I^2 >80%.

Re-analysis of studies reporting person-level outcome measurements

Because some of the 14 studies providing analyzable outcome data had multiple outcomes (or multiple outcomes scales) there were a total of 27 comparisons with analyzable data. (Appendix Figures 17-42 displays graphs of the person level outcome results.) Average

fixed effect estimates for each analysis are shown in Table 5; random effects estimates were generally similar (Appendix Table 6). In eight of the 14 studies (57%) (17 of the 27 unique comparisons [63%]), there was statistically significant person-level HTE for at least one outcome. Again, the variation in individual effects was often large compared to the average effect. However, given the lower number of participants per study and periods per participant and also different analytic approach, estimates of I^2 were much less precise in these studies.

Discussion

This review documents that multi-person N-of-1 studies rarely examine HTE. Only 8% of 62 multi-person N-of-1 studies described statistical tests examining HTE, but these generally involved comparisons of treatment effects among groups of patients (e.g., based on age or sex) rather than across individuals. Only two studies in the whole of the literature tested for person-level HTE.[15;16] Nevertheless, analyzable person-level results are sometimes reported in multi-person N-of-1 studies, as outcomes or as treatment effects, suitable for the analysis of person-level HTE. Our re-analyses of the totality of available data from these studies (n=25) suggested the presence of substantial non-random variation in treatment effects across individuals in most studies. This was evident when considering statistical tests for the variation of treatment effects among patients and also by qualitative assessment of the magnitude of effect variation. This represents the first broad empirical examination with re-analysis of person-level HTE across multi-person N-of-1 studies, and it provides some general support for the *a priori* assumption of individual patient variation in treatment response that broadly motivates personalized medicine.

In contrast to parallel-group studies that establish efficacy in a group of patients with a common condition, N-of-1 studies establish the effects of an intervention in an individual.[17] In

this respect, N-of-1 studies can be thought of as adjuncts to clinical care, where the goal is to select the right treatment for a particular patient, rather than as a research tool, where the goal is to create new generalizable knowledge.[18;19] Indeed, the results of traditional N-of-1 studies may be generalizable only to the future treatment response of the patient in the trial, not to other patients. Nevertheless, using Bayesian meta-analytic techniques, Zucker et al. showed how the average treatment effect at the population-level can also be estimated by combining multi-person N-of-1 studies testing similar interventions in similar patients with the same outcome measures.[14] Similar Bayesian methods have also been suggested for analysis of group-level HTE.[20]

Herein, we demonstrate yet a new application of N-of-1 studies, to explore person-level HTE. This application has important research and clinical implications, even when the determinants of HTE remain unidentified. It is particularly of interest that there was apparent variation in the *degree* of person-level HTE found across conditions and treatments. Since the degree of variation across individuals sets the upper bound for the amount of HTE that might be explainable by observable characteristics, such as clinical or genomic variables, searching for subgroup effects in the absence of person-level HTE is a futile exercise.[4;21]

An interesting example of how person-level HTE can vary across different conditions comes from the study of Johannessen et al **(Figure 3)**.[15] These investigators conducted N-of-1 patient studies comparing cimetidine to placebo for patients presenting with dyspeptic symptoms and reported person-level effects by subgroups of disease categories. Among 46 trial completers, cimetidine had a significant effect for most patients (57%), as it did at the aggregate level. However, not only was there substantial person-level HTE, but person-level HTE varied across conditions, being much more pronounced in non-ulcer dyspepsia (I^2 = 75%) compared to peptic

ulcer disease (I^2 = 35%) (Figure 3)— despite the very similar overall effects seen in these two conditions.

Finding variation in person-level response in multi-person N-of-1 studies identifies those conditions for which N-of-1 studies are likely to be clinically relevant. For condition-treatment combinations shown to have low person-level HTE, single subject studies are highly unlikely to be clinically informative, and the average results from trials (i.e., "one-size-fits-all" effects) are more apt to be applicable to individuals.[22;23] On the other hand, N-of-1 studies may be highly clinically informative for condition-treatments with a high degree of person-level HTE. These conditions would also be potentially higher yield for examining predictors of HTE (genomic or otherwise).

Our findings also have implications for clinical practice and formulary design. For conditions marked by high person-level HTE, even when trials show that one treatment is better on average than others, having a variety of medication options would be useful to optimize outcomes across all patients, particularly for chronic conditions such as those studied here where empiric trials of alternative medications to find the best treatment for an individual might be feasible. For example, the study by March et al. shows that while patients with osteoarthritis on average had less pain and less stiffness with diclofenac, some patients had improved symptoms on paracetemol.[24] This person-level heterogeneity of treatment effect may not be detectable in conventional parallel arm trials employing conventional subgroup analysis.[21]

While more studies combining N-of-1 studies are needed to understand the extent of person-level HTE, future studies need to apply greater methodological rigor to improve the state-of-the-science on evaluation of individual treatment effects.[25] While the recently published CONSORT Extension for N-of-1 Trials may help improve reporting, a tabulation of all

information (possibly electronically available) appears the most straightforward way to facilitate the clinical interpretation of these studies.[26] Such reporting allows the inspection of trajectories over time and may reveal patterns that are not captured by regression models. Complete reporting would also facilitate the development and evaluation of methods for the analysis of single subject experiments, particularly its use to better understand the extent and importance of person-level HTE.

The limitations of this review reflect, to a large extent, the limitations of the data in primary studies. Many conditions are not amenable to the N-of-1 design (e.g. because treatment effects are cumulative or because outcomes are observed only once). Further, even for conditions and treatment that are potentially amenable to this design, many important disease categories lacked published N-of-1 studies. We relied on published studies only and our analytic cohort may be an underestimation of the true prevalence of these studies—particularly for N-of-1studies, which may frequently be conducted without the intention of future publication.

In addition, our conclusions regarding the ubiquity of HTE in the data we reanalyzed should be interpreted in the context of several important limitations. First, there were only a limited number of available studies that reported data sufficient to analyze, and therefore we present only a very partial picture of the full scope of inter-individual variation in effects across clinical conditions. Furthermore, among the studies that did have data, only fairly small numbers of patients were observed over a small number of treatment periods and we frequently had to rely on data summaries provided by the authors (e.g., person-level treatment effects and their sampling variance); these data limitations precluded the use of more complex models, for example models that account for period effects or other effects of time on the outcome.[3]

Our review has demonstrated that HTE remains almost totally unexplored in multi-person N-of-1 studies, which are uniquely capable of exploring variations in individual (person-level) treatment effects. Our re-analysis of the data from these studies represents the first systematic attempt to obtain empirical support for the *a priori* argument that treatment effects vary across individual patients, an assumption which underpins all efforts to personalize treatment selection. In this sample, person-level HTE appears to be common and large enough to be clinically meaningful; the degree of person-level HTE appears to vary across conditions and outcomes. Thus, multi-person N-of-1 studies are an under-utilized tool to identify where person-level HTE may be substantial, and where efforts to find molecular or clinical predictors of response heterogeneity should be focused. In such conditions, parallel arm studies might yield results that are over-generalized for patient level decision making.

Funding Source

This work was supported by the National Pharmaceutical Council. Additional support was provided by the Patient-Centered Outcomes Research Institute (PCORI) Award (Predictive Analytics Resource Center (SA.Tufts.PARC.OCSO.2018.01.25), and the National Institutes of Health (3UL1TR001079-04S1).

Disclosures:

All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors or Methodology Committee.

Acknowledgement

We would like to acknowledge Issa Dahabreh, MD, MS, Assistant Professor of Health Services, Policy and Practice, Assistant Professor of Epidemiology, Brown University, for statistical advice.

We would like to acknowledge Tatum Williamson, MS, Research Assistant, Predictive Analytics and Comparative Effectiveness Center, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, for assistance with updating literature.

Contributorship statement

DK, GR made substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work. DK, GR are responsible for drafting the work or revising it critically for important intellectual content. DK, GR, EB, LL, JS, JC, JL,

RD, RK have given final approval of the version to be published. DK, GR have made an agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Competing interests

None declared.

Data sharing statement

No additional data are available.

Reference List

(1) Wallach JD, Sullivan PG, Trepanowski JF, Steyerberg EW, Ioannidis JP. Sex based subgroup differences in randomized controlled trials: empirical evidence from Cochrane meta-analyses. *BMJ* 2016; 355:i5826.

(2) Hertz DL, McLeod HL. Use of pharmacogenetics for predicting cancer prognosis and treatment exposure, response and toxicity. *J Hum Genet* 2013; 58(6):346-352.

(3) Kitsios GD, Kent DM. Personalised medicine: not just in our genes. *BMJ* 2012; 344:e2161.

(4) Senn S. Individual response to treatment: is it a valid assumption? *BMJ* 2004; 329(7472):966-968.

(5) Kalow W, Tang BK, Endrenyi L. Hypothesis: comparisons of inter- and intra-individual variations can substitute for twin studies in drug research. *Pharmacogenetics* 1998; 8(4):283-289.

(6) Gabler NB, Duan N, Vohra S, Kravitz RL. N-of-1 trials in the medical literature: a systematic review. *Med Care* 2011; 49(8):761-768.

(7) Finding What Works in Health Care: Standards for Systematic Reviews.Institute of Medicine (US) Committee on Standards for Systematic Reviews of Comparative Effectiveness Research. 2011.

(8) Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009; 151(4):264-9, W64.

(9) Methods Guide for Effectiveness and Comparative Effectiveness Review. AHRQ Publication No 10(11)-EHC063-EF Chapters 2011 Available from: URL:http://effectivehealthcare.ahrq.gov/

(10) Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011; 343:d5928.

(11) DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7(3):177-188.

(12) Schmidt FL, Oh IS, Hayes TL. Fixed- versus random-effects models in meta-analysis: model properties and an empirical comparison of differences in results. *Br J Math Stat Psychol* 2009; 62(Pt 1):97-128.

(13) Cochran WG. The combination of estimates from different experiments. Biometrics 10, 101-129. 1954.

(14) Zucker DR, Ruthazer R, Schmid CH. Individual (N-of-1) trials can be combined to give population comparative treatment effect estimates: methodologic considerations. *J Clin Epidemiol* 2010; 63(12):1312-1323.

(15) Johannessen T, Petersen H, Kristensen P, Fosstvedt D, Kleveland PM, Dybdahl J et al. Cimetidine on-demand in dyspepsia. Experience with randomized controlled single-subject trials. *Scand J Gastroenterol* 1992; 27(3):189-195.

(16) Pereira JA, Holbrook AM, Dolovich L, Goldsmith C, Thabane L, Douketis JD et al. Are brand-name and generic warfarin interchangeable? Multiple n-of-1 randomized, crossover trials. *Annals of Pharmacotherapy* 2005; 39(7-8):1188-1193.

(17) Guyatt GH, Heyting A, Jaeschke R, Keller J, Adachi JD, Roberts RS. N of 1 randomized trials for investigating new drugs. *Control Clin Trials* 1990; 11(2):88-100.

(18) Guyatt G, Sackett D, Taylor DW, Chong J, Roberts R, Pugsley S. Determining optimal therapy--randomized trials in individual patients. *N Engl J Med* 1986; 314(14):889-892.

(19) Guyatt G, Sackett D, Adachi J, Roberts R, Chong J, Rosenbloom D et al. A clinician's guide for conducting randomized trials in individual patients. *CMAJ* 1988; 139(6):497-503.

(20) Henderson NC, Louis TA, Wang C, Varadhan R. Bayesian analysis of heterogeneous treatment effects for patient-centered outcomes research. *Health Serv Outcomes Res Methodol* 2016; 16(4):213-233.

(21) Dahabreh IJ, Hayward R, Kent DM. Using group data to treat individuals: understanding heterogeneous treatment effects in the age of precision medicine and patient-centred evidence. *Int J Epidemiol* 2016.

(22) Simon G. Choosing a first-line antidepressant: equal on average does not mean equal for everyone. *JAMA* 2001; 286(23):3003-3004.

(23) Simon GE, Psaty BM, Hrachovec JB, Mora M. Principles for evidence-based drug formulary policy. *J Gen Intern Med* 2005; 20(10):964-968.

(24) March L, Irwig L, Schwarz J, Simpson J, Chock C, Brooks P. n of 1 trials comparing a non-steroidal anti-inflammatory drug with paracetamol in osteoarthritis. *BMJ* 1045; 309(6961):1041-1045.

(25) Kravitz RL, Duan N. eds, and the DEcIDE Methods Center N-of-1 Guidance Panel. Design and Implementation of N-of-1 Trials: A User's Guide. AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality. 2014.

(26) Vohra S, Shamseer L, Sampson M, Bukutu C, Schmid CH, Tate R et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. *BMJ* 2015; 350:h1738.

Table 1. Evidence Map of Multi-person N-of-1and Repeated Period Crossover Studies

Description	Multi-person N-of-1 Studies (n=62)
Publication Years	1979-2017
Subjects	**Total N (median, IQR)**
Enrolled	2153 (16, 9-42)
Completed	1705 (12, 7-32)
Intervention & Comparisons	
Head-to-head active drugs	10
Placebo	47
Active drug and placebo	1
Population	
Pediatric	12
Adults	50
Major Systems Studied	
Arthritis/Rheumatology	10
Cardiovascular	3
Gastrointestinal	7
Hypertension	1
Psychiatry	9
Neurology	16
Respiratory	9
Miscellaneous*	7
Top 5 Disease Conditions	
ADHD	6
Angina	3
Chronic Pain	5
GERD	5
Obstructive Airway	6
Osteoarthritis	6

**Sleep disorders, Allergy, Cancer, Muscular, Vascular (for multi-person N-of-1); Pain, Urology, GYN, , Heme/Onc, Allergy, Dermatology, Drug abuse, Endocrine, Lipids, Nephrology, Ophthalmology, Respiratory (for Repeated Cross-over Studies). ADHD, Attention-deficit hyperactivity disorder; GERD, Gastro-esophageal regurgitation disorder; IQR, Interquartile range; n, number of participants*

Table 2. Survey of HTE Assessment in Multi-person N-of-1 Studies

HTE Reporting	Multi-person N-of-1 Studies (n=62)
Qualitative description	92%
Person-level outcomes	52%
Person-level treatment effects	32%
Statistical analysis of person-level effects (e.g. p-values)	21%
Any statistical test for HTE	8%*
Claims of heterogeneity	15%

* Only 2 studies reported person-level HTE, the remaining 3 studies reported group level effect.

Table 3. Characteristics of studies reporting person-level data

Author, Year	Disease	Number enrolled (analyzed)	Intervention	Comparator	Cross-over periods	Total intervention duration	Outcome measures per period
Studies with re-analyzable person-level outcomes							
Camfield, 1996	Mental retardation with fragmented sleep	6 (6)	Melatonin	Placebo	7	10 wk	14
Hinderer, 1990	Traumatic spinal cord injury	5 (5)	Baclofen	Placebo	3	9 wk	2
Langer, 1993	Gastroesophageal reflux	2 (2)	Cisapride	Placebo	3	6 wk	5
Lashner, 1990	Ulcerative colitis	7 (6)	Nicotine	Placebo	4	8 wk	1
Maier, 1994	Chronic depression	10 (9)	Sulpiride	Placebo	4	28 wk	42
Mandelcorn, 2004	Brain injury	4 (4)	Ondansetron	Placebo	4	5 wk	1
McQuay, 1994	Neuropathic pain	19 (19)	Dextromethorphan	Placebo	5	20 d	1
Miyazaki, 1995	Unstable angina	22 (22)	Isosorbide dinitrate	Isosorbide dinitrate: intermittent injection	3	9 d	6
Nathan, 2006	Pediatric brain tumor	12 (7)	Ondansetron & metopimazine	Ondansetron & placebo	Unclear	189 d	unclear
Parodi, 1979	Unstable angina	12 (12)	Verapamil	Placebo	4	10 d	unclear
Parodi, 1986	Unstable angina	10 (10)	Verapamil	Propranolol, placebo	8	18 d	unclear
Tison, 2012	Levodopa-induced dyskinesia in Parkinson's disease patients	10 (10)	Simvastatin	Placebo	6	96 d	1
Studies with re-analyzable person-level treatment effects							
Emmanuel, 2012	Chronic intestinal pseudo-obstruction	7 (4)	Prucalopride	Placebo	16	48 wk	21
Haas, 2004	Chronic tension-type and migraine headache	39 (16)	Dextroamphetamine	Equi-stimulatory caffeine	8	20 d	20
Jaeschke, 1991	Fibromyalgia	22 (23)	Amitriptyline	Placebo	6	12 wk	2
Johannessen, 1992	Dyspepsia	68 (46)	Cimetidine	Placebo	12	184 d	15
Lipka, 2017	Autoimmune myasthenia	4 (4)	Ephedrine	Placebo	4	6 wk	1

	gravis						
Mahon, 1996	Irreversible chronic airflow limitation	16 (14)	Theophylline	Placebo	8	73 d	1
March, 1994	Osteoarthritis	25 (15)	Diclofenac	Paracetamol	6	12 wk	14
Patel, 1991	Nonreversible chronic airflow limitation	26 (18)	Ipratropium bromide / theophylline / salbutamol/ beclomethasone	Placebo	6	6 wk	Unclear
Wallace, 1994	Attention deficit hyperactivity disorder	11 (7)	Methylphenidate	Placebo	14	14 d	1
Woodfield, 2005	Skeletal muscle cramps	13	Quinine	Placebo	6	14 wk	2
Zucker, 2006	Fibromyalgia	58	Amitriptyline and Placebo	Amitriptyline and fluoxetine combination	6	36 wk	1
Study with both person-level data							
Pereira, 1995	Atrial fibrillation / deep venous thrombosis	7	Generic warfarin	Coumadin	10	30 wk	2
Joy, 2014	Statin-related myalgia	8 (7)	Statin	Placebo	6	33 wk	3
Study with insufficiently reported person-level data							
Person-level outcome data							
Denburg, 1994	Systemic lupus erythematosus	10	Prednisone	Placebo	6	30 wk	1
Mitchel, 2015	Fatigue in advanced cancer	43 (33)	Methylphenidate	Placebo	6	18 d	6
Nikles, 2000	Osteoarthritis	14	Ibuprofen	Paracetamol; Placebo	6	12 wk	14
Nikles, 2015	Dry mouth in advanced cancer	17 (4)	Pilocarpine	Placebo	6	18 d	6
Nikles, 2017	Acquired brain injury	53 (38)	Nervous system stimulants	Placebo	6	18 d	6
Reitberg, 2002	Allergic rhinitis	36	Loratadine and chlorpheniramine maleate	loratadine with placebo	8	32 d	4
Sheather-Reid, 1998	Chronic pain	8	Ibuprofen / Codeine	Placebo	6	12 wk	14
Person-level treatment effects							

Huber, 2007	Juvenile idiopathic arthritis	6	Amitriptyline	Placebo	6	17 wk	12
Privitera, 1994	Partial seizure	16	Dezinamide	Placebo	6	35 wk	6
Wegman, 2003	Osteoarthritis	13	Paracetamol	NSAIDs	10	20 wk	14
Wegman, 2005	Regular Temazepam users	15	Temazepam	Placebo	10	10 wk	7

Table 4. Analysis results of studies reporting person-level treatment effects

			Main Effect	*Person-Level Heterogeneity of Treatment Effect*		
Author Year	**Outcome**	**Range of the scales (severity)**	**Treatment effect (CI)**	**P for HTE***	**Treatment Effect Range**	**I-square % (CI)**
Emmanuel 2012	Bloating	0-4 (0=absent to 4=worst)	-0.344 (-0.619 to -0.069)	<0.001	-1.1 to -0.1	94 (88 to 97)
	Pain	0-4 (0=absent to 4=worst)	-0.440 (-0.771 to -0.110)	<0.001	-0.2 to -1.4	96 (92 to 98)
Haas 2004	Chronic tension-type headache grade	0-3 (0=none to 3=severe)	0.772 (0.454 to 1.090)	<0.001	0.04 to 1.9	84 (76 to 90)
	Chronic migraine headache grade	0-3 (0=none to 3=severe)	0.542 (0.354 to 0.731)	0.067	0.2 to 0.83	37 (0 to 65)
Jaeschke 1991	7-point symptom scale	1-7 (higher scores represent better function)	0.427 (0.210 to 0.645)	<0.001	-1.02 to 3.18	85 (79 to 89)
	Tender point changes count	Number of tender points	1.320 (0.404 to 2.236)	<0.001	-4.33 to 9.0	72 (57 to 82)
Johannessen 1992	6-point symptom scale	0-6 (0=NR to 6=NR)	0.698 (0.466 to 0.931)	<0.001	-1.67 to 3.17	66 (53 to 75)
Joy 2014	VAS myalgia Score	0-100mm (0=none to 100=worst)	0.119 (-2.283 to 2.521)	0.996	-8,10 to 9.45	0 (0 to 68)
	Symptom-specific VAS	0-100mm (0=none to 100=worst)	1.937 (0.179 to 3.696)	0.797	-8.0 to 18.05	0 (0 to 68)
	Pain severity score	0-10 (0=none to 10=worst)	0.086 (-0.215 to 0.387)	0.986	0.0 to 1.0	0 (0 to 68)
	Pain interference score	0-10 (0=none to 10=worst)	-0.016 (-0.095 to 0.064)	0.917	-0.02 to 0.75	0 (0 to 68)
Lipka 2017	Quantitative myasthenia gravis score	0-3 (0=none to 3=severe)	1.006 (0.215 to 1.797)	0.803	0.67 to 1.67	0 (0 to 85)
	Myasthenia gravis composite	0-50	2.891 (0.348 to 5.433)	0.177	-1.05 to 5.12	39 (0 to 80)
	MG-ADL	0-24	1.099 (-0.277 to 2.474)	0.047	0.03 to 3.0	62 (0 to 87)
	VAS score	0-10 (0=none to 100=worst)	1.275 (-0.115 to 2.665)	0.190	-0.01 to 3.02	37 (0 to 78)
Mahon 1996	Dyspnea in likert Scale	1-7 (1=extremely short of breath to 7=no shortness)	0.125 (-0.181 to 0.430)	<0.001	-0.57 to 0.89	78 (58 to 88)
March 1994	Mean pain score on VAS	5 point Likert scale (0-100mm)	-7.093 (-11.939 to -2.248)	<0.001	-33.8 to 4.1	98 (97 to 98)
	Mean stiffness score on VAS	5 point Likert scale (0-100mm)	-5.992 (-11.280 to -0.704)	<0.001	-36 to 10.7	97 (96 to 98)
Patel 1991**	4-item symptom questionnaire (All compared to placebo)	1-7 (1=extremely short of breath to 7=no shortness of breath)	0.340 (0.253 to 0.422)	<0.001	-0.34 to 3.1	91 (87 to 94)
	4-item symptom questionnaire (use of ipratropium bromide)		0.675 (0.264 to 1.085)	<0.001	-0.22 to 3.1	87 (78 to 92)
	4-item symptom questionnaire (use of salbutamol)		0.865 (0.042 to 1.687)	<0.001	0.46 to 1.3	94 (NA)
	4-item symptom questionnaire (use of theophylline)		0.025 (-0.434 to 0.484)	0.172	-0.34 to 0.18	30 (0 to 93)
Pereira	INR (diff)	Target INR range of 2.0–3.0	0.027 (-0.155 to 0.209)	0.477	-0.28 to 0.37	0 (0 to 75)

1995						
Wallace 1994	Conners 15-item rating scale scores	0-3 (NR)	0.759 (0.341 to 1.178)	0.747	0.42 to 1.22	0 (0 to 79)
Woodfield 2005	Changes in number of cramps	Number – mean difference	-18.823 (-28.527 to -9.120)	<0.001	-77 to -2	92 (87 to 95)
	Total days with cramps	days	-6.181 (-9.798 to -2.563)	<0.001	-13 to -1	94 (90 to 96)
Zucker 2006	FIQ	0-100 (0=best to 100=worst)	-5.019 (-8.784 to -1.254)	0.999	-32.0 to 0.98	0 (0 to 37)

* The significance of person-level HTE was assessed by Cochran's chi-square-based test
** One subject had beclomethasone

Table 5. Studies reporting person-level outcomes

Author Year	Outcome	Definition / Range of the Scales (severity)	*Main Effect* Fixed Treatment Effect	*Person-level Heterogeneity of Treatment Effect* P for Person Treatment Interaction*	Treatment Effect Range Lower Range (CI) Upper Range (CI)	I-square % (CI)
Camfield 1996	Nights without awakening	Between 10:00 PM and 7:00 AM per day	0.865 (0.215 to 1.516)	0.456	0.12 to 2.0	0 (0 to 79)
Hinderer 1990	Anxiety	Beck Inventory-A anxiety scale 0-3 (0 = never, 3 = almost all the time)	0.000 (0.000 to 0.000)	<0.001	-6.38 to 0.000	91 (81 to 95)
Joy 2014	Myalgia score	Visual Analogue Score for myalgia (0=none to 100=worst)	3.3812 (-2.668 to 9.430)	0.565	-11.66 to 60.79	0 (0 to 68)
Langer 1993	Vomiting	Number of episodes	-1.204 (-2.494 to 0.086)	0.136	-1.34 to 0.17	87 (NA)*
Lashner 1990	Symptom score: abdominal pain	Symptom scores 0-100 (0 = best, 100 = worst)	-3.615 (-16.982 to 9.751)	0.007	-35.0 to 15.0	37 (0 to 73)
	Symptom score: bowel movements/day		-0.538 (-1.215 to 0.138)	0.001	-3.0 to 1.0	56.6 (0 to 81)
	Symptom score: consistency of bowel movements		7.000 (-7.551 to 21.551)	0.013	-25.5 to 33.0	28 (0 to 69)
	Symptom score: hematochezia		2.308 (-17.210 to 21.826)	0.003	-38.0 to 47.5	47 (0 to 78)
	Symptom score: general sense of well-being		-6.538 (-25.352 to 12.275)	0.008	-43.0 to 35.0	35 (0 to 73)
Maier 1994	SCL-90 subscales: Depressed mood	Self-rating inventory to measure the effects of drug	-3.536 (-6.718 to -0.354)	<0.001	-17.8 to 2.74	58 (12 to 80)
	SCL-90 subscales: Anxiety		-3.753 (-6.582 to -0.924)	<0.001	-17.4 to 2.5	66 (30 to 83)
	SCL-90 subscales: Somatization		-1.419 (-4.316 to 1.478)	0.869	-6.0 to 2.7	0 (0 to 65)
Mandelcorn 2004	Self-Assessment score	0–5 (0 = worst, 5 = best)	-2.052 (-8.865 to 4.761)	0.05	-7.7 to 4.9	0 (0 to 85)
	Lower extremity ataxia	Fugl-Meyer: 3-point (0 cannot be performed to 2 can be fully performed)	12.494 (-3.155 to 28.142)	0.025	-6.42 to 36.76	35 (0 to 77)
	Truncal ataxia	AMTI forceplate®: NR *Berg Balance Scale® 0–56, with a higher score indicating a better performance*	1.196 (-2.866 to 5.257)	0.690	-0.52 to 2.20	0 (0 to 85)
	Upper extremity ataxia	Purdue Pegboard Test®: pegs inserted into the board with each hand in 30 sec	-0.498 (-3.546 to 2.550)	0.382	-3.68 to 1.42	0 (0 to 85)

Author Year	Outcome	Definition / Range of the Scales (severity)	*Main Effect*	*Person-level Heterogeneity of Treatment Effect*		
			Fixed Treatment Effect	P for Person Treatment Interaction*	Treatment Effect Range Lower Range (CI) Upper Range (CI)	I-square % (CI)
		Minnesota Placing Test®: reach out, grasp, and place blocks in a specific order				
McQuay 1994	VAS Pain Intensity	0-100 (0 = no pain, 100 = worst possible pain)	-1.094 (-5.572 to 3.383)	0.004	-8.0 to 10.1	0 (0 to 49)
	VAS Relief Intensity	0-100 (0 = no relief, 100 = complete pain relief)	-3.913 (-11.729 to 3.903)	0.038	-28.4 to 5.15	0 (0 to 49)
Miyazaki 1995	Incidence of angina	Either ST-segment elevation or depression at rest	0.496 (-0.206 to 1.199)	0.125	-16.19 to 17.11	0 (0 to 60)
Nathan 2006	Emetic episodes per day	complete response (0 episodes/day), major response (1–2 episodes/day), or failure (>2 episodes/day)	-0.095 (-0.514 to 0.325)	0.001	-16.5 to 2.08	59 (6 to 82)
Parodi 1979	Ischemic attacks	ST elevation or depression (details NR)	-1.544 (-1.838 to -1.251)	0.007	-16.21 to -0.34	48 (0 to 73)
Parodi 1986	Asymptomatic ST elevation (After verapamil)	0.1 mV of ST-segment elevation measured 20 ms after the J point	-1.637 (-1.994 to -1.279)	0.110	-2.37 to -1.30	6 (0 to 65)
	Asymptomatic ST depression (After verapamil)	More than 0.2 mV of ST-segment depression measured 80 ms after the J point	-1.083 (-1.903 to -0.262)	0.401	-17.42 to -0.90	0 (0 to 62)
	Symptomatic ST elevation (After verapamil)		-1.580 (-1.906 to -1.254)	<0.001	-15.40 to -1.45	0 (0 to 62)
	Symptomatic ST Depression (After verapamil)		-0.990 (-1.411 to -0.569)	0.002	-2.53 to -0.52	6 (0 to 64)
	Asymptomatic ST elevation (After propranolol)		0.100 (-0.086 to 0.286)	0.006	-0.77 to 1.38	62 (25 to 81)
	Asymptomatic ST depression (After propranolol)		0.339 (-0.168 to 0.845)	0.964	-18.3 to 0.83	0 (0 to 62)
	Symptomatic ST elevation (After propranolol)		-0.002 (-0.177 to 0.173)	0.063	-14.9 to 0.68	46 (0 to 74)
	Symptomatic ST Depression (After propranolol)		-0.374 (-0.709 to -0.039)	0.023	-17.1 to -0.73	4 (0 to 64)
Pereira 1995	INR	Target INR range of 2.0–3.0	-0.126 (-0.312 to 0.060)	0.433	-0.42 to 0.16	0 (0 to 71)
Tison 2012	Troublesome dyskinesia	7 points scale (1 = extremely uncomfortable, 7 = not at all uncomfortable)	0.167 (-0.449 to 0.783)	0.593	-0.67 to 1.83	0 (0 to 62)

* The significance of person-level HTE was assessed by a likelihood ratio test comparing the two models – model with common treatment effect and model with treatment-by-participant interactions

Figure Legend

Figure 1: The Figure provides a schematic description of: person-level outcomes (outcomes for each patient during each treatment period); person-level effects (contrasts of the outcomes for each patient in one treatment condition *versus* another); and person-HTE (between patient contrasts of effects).

Figure 2. Study Flow Diagram represents the flow of eligible studies included in this review

Figure 3. Person-level variation across different disease conditions. This figure depicts the results of 46 different N-of-1 trials of cimetidine as reported by Johanessen et al [12]. The effect of cimetidine versus placebo was measured in each subject across 12 cross-over periods over the span of 184 days. While cimetidine had a similar average effect regardless of the index condition, there was far greater consistency of effect in patients with peptic ulcer disease and much more variation in effect among patients with non-ulcer dyspepsia.

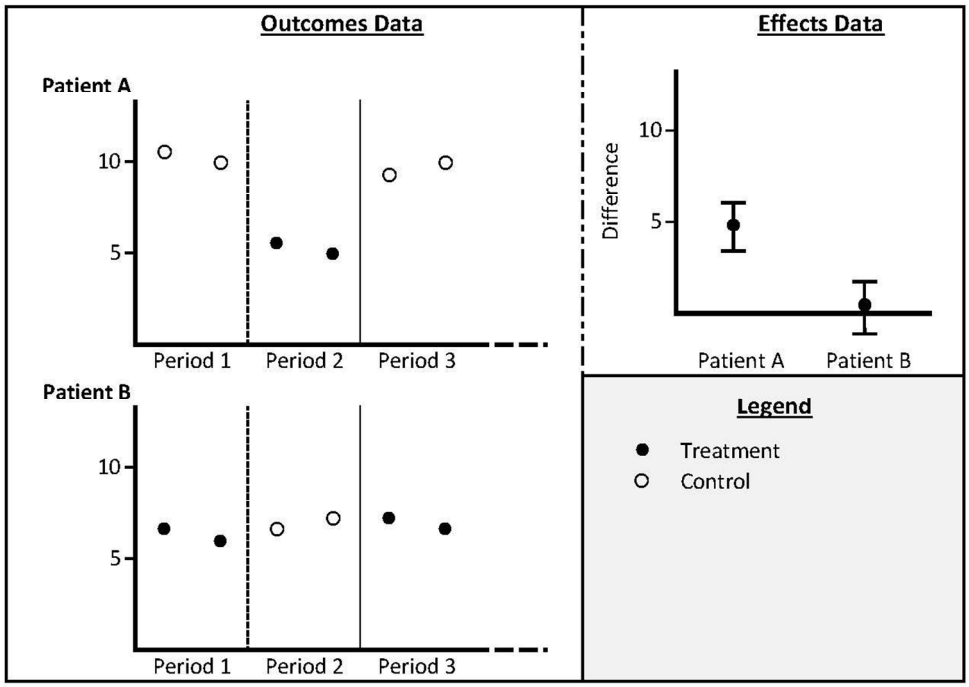


Figure 1: The Figure provides a schematic description of: person-level outcomes (outcomes for each patient during each treatment period); person-level effects (contrasts of the outcomes for each patient in one treatment condition versus another); and person-HTE (between patient contrasts of effects).

111x79mm (300 x 300 DPI)

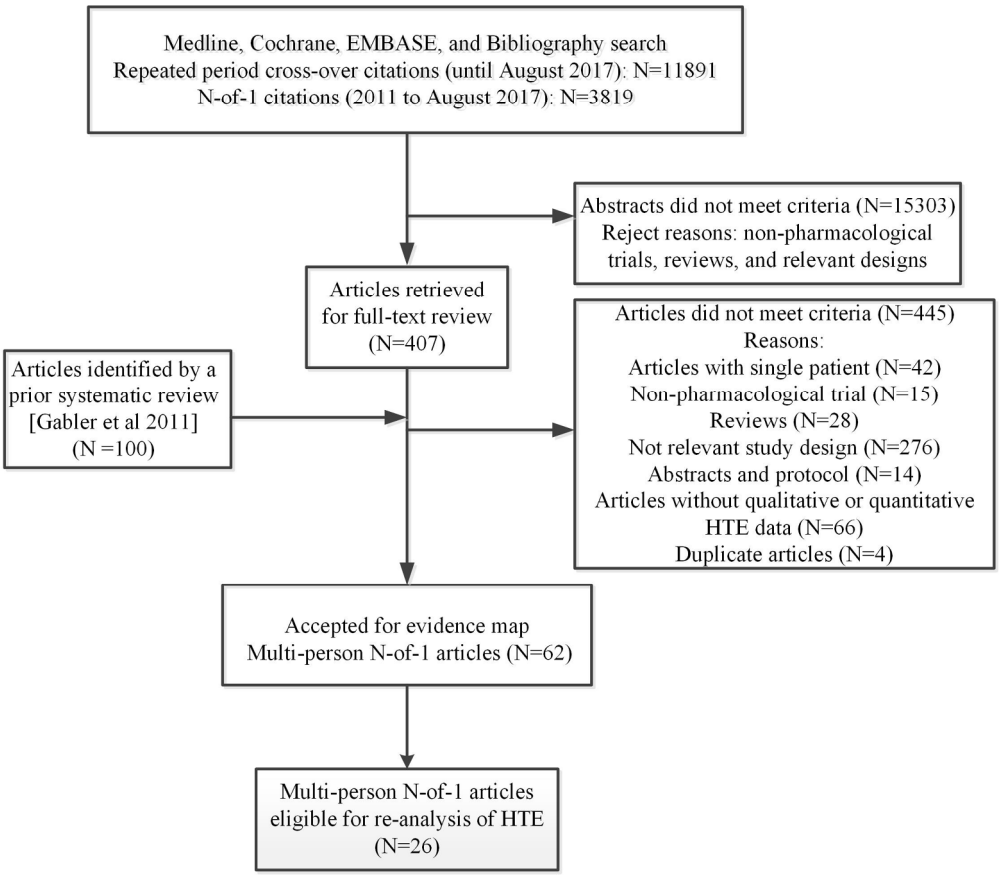


Figure 2. Study Flow Diagram represents the flow of eligible studies included in this review

189x164mm (300 x 300 DPI)

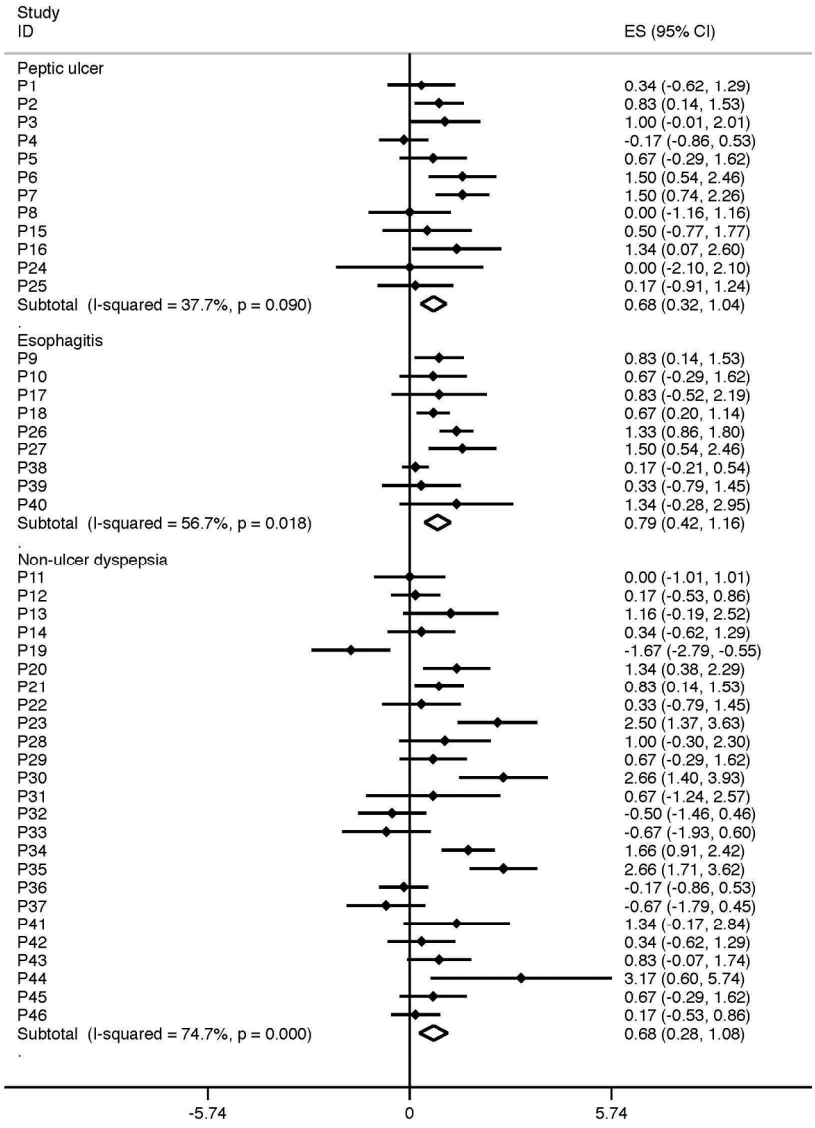


Figure 3. Person-level variation across different disease conditions. This figure depicts the results of 46 different N-of-1 trials of cimetidine as reported by Johanessen et al 12. The effect of cimetidine versus placebo was measured in each subject across 12 cross-over periods over the span of 184 days. While cimetidine had a similar average effect regardless of the index condition, there was far greater consistency of effect in patients with peptic ulcer disease and much more variation in effect among patients with non-ulcer dyspepsia.

150x203mm (300 x 300 DPI)

Appendix Materials

Appendix Table 1: N-of-1 Trial Searches

1.	randomized controlled trial.pt.
2.	controlled clinical trial.pt.
3.	randomized controlled trials/
4.	Double-blind Method/
5.	Single-Blind Method/
6.	clinical trial.pt.
7.	Clinical Trials.mp. or exp Clinical Trials/
8.	random$.tw.
9.	trial$.tw.
10.	Cross-Over Studies/
11.	or/1-10
12.	n-of-1.af.
13.	11 and 12
14.	(single-subject or single-patient or single case or single-case or within-patient).af.
15.	((single adj1 patient) or (single adj1 subject)).tw.
16.	14 or 15
17.	12 and 16
18.	multi-crossover.mp.
19.	12 and 18
20.	13 or 17 or 19
21.	limit 19 to yr="2010 - 2017"

Appendix Table 2: Repeated Period Crossover Trials

1.	(repeat$ or rotat$).af.
2.	((three or four or five or six) and period).tw.
3.	(multi- or multiple).tw.
4.	(three-period or four-period or five-period or six-period).tw.
5.	(three-way or four-way or five-way or six-way).tw.
6.	or/1-5
7.	Cross-Over Studies/ or (cross-over or crossover).af.
8.	6 and 7
9.	randomized controlled trial.pt.
10.	controlled clinical trial.pt.
11.	randomized controlled trials/
12.	Double-blind Method/
13.	Single-Blind Method/
14.	clinical trial.pt.
15.	Clinical Trials.mp. or exp Clinical Trials/
16.	random$.tw.
17.	trial$.tw.
18.	or/9-17
19.	8 and 18
20.	(dt or de or tu).fs.
21.	19 and 20
22.	7 and 20
23.	"Reproducibility of Results"/
24.	16 and 22
25.	limit 22 to english language
26.	9 or 10 or 11 or 14 or 15 or 16
27.	7 or 23
28.	20 and 26 and 27
29.	random.af.
30.	9 or 10 or 11 or 14 or 15 or 29
31.	ae.fs.
32.	20 or 31
33.	27 and 30 and 32
34.	limit 33 to (english language and humans)
35.	periods.af.

36.	6 or 35
37.	33 and 36
38.	Animals/ not human/
39.	37 not 38

Appendix Table 3: Reference List of Included Studies

1.	Nikles CJ, McKinlay L, Mitchell GK, Carmont SA, Senior HE, Waugh MC et al. Aggregated n-of-1 trials of central nervous system stimulants versus placebo for paediatric traumatic brain injury--a pilot study. Trials [Electronic Resource] 2014; 15:54.
2.	Tison F, Negre-Pages L, Meissner WG, Dupouy S, Li Q, Thiolat ML et al. Simvastatin decreases levodopa-induced dyskinesia in monkeys, but not in a randomized, placebo-controlled, multiple cross-over ("n-of-1") exploratory trial of simvastatin against levodopa-induced dyskinesia in Parkinson's disease patients. Parkinsonism & Related Disorders 2013; 19(4):416-421.
3.	Rascol O, Ferreira J, Negre-Pages L, Perez-Lloret S, Lacomblez L, Galitzky M et al. A proof-of-concept, randomized, placebo-controlled, multiple cross-overs (n-of-1) study of naftazone in Parkinson's disease. Fundamental & Clinical Pharmacology 2012; 26(4):557-564.
4.	Emmanuel AV, Kamm MA, Roy AJ, Kerstens R, Vandeplassche L. Randomised clinical trial: the efficacy of prucalopride in patients with chronic intestinal pseudo-obstruction--a double-blind, placebo-controlled, cross-over, multiple n = 1 study. Alimentary Pharmacology & Therapeutics 2012; 35(1):48-55.
5.	Yelland MJ, Poulos CJ, Pillans PI, Bashford GM, Nikles CJ, Sturtevant JM et al. N-of-1 randomized trials to assess the efficacy of gabapentin for chronic neuropathic pain. Pain Medicine 2009; 10(4):754-761.
6.	Nonoyama ML, Brooks D, Guyatt GH, Goldstein RS. Effect of oxygen on health quality of life in patients with chronic obstructive pulmonary disease with transient exertional hypoxemia. American Journal of Respiratory & Critical Care Medicine 2007; 176(4):343-349.
7.	Huber AM, Tomlinson GA, Koren G, Feldman BM. Amitriptyline to relieve pain in juvenile idiopathic arthritis: a pilot study using Bayesian metaanalysis of multiple N-of-1 clinical trials. Journal of Rheumatology 2007; 34(5):1125-1132.
8.	Yelland MJ, Nikles CJ, McNairn N, Del Mar CB, Schluter PJ, Brown RM. Celecoxib compared with sustained-release paracetamol for osteoarthritis: a series of n-of-1 trials. Rheumatology 2007; 46(1):135-140.
9.	Zucker DR, Ruthazer R, Schmid CH, Feuer JM, Fischer PA, Kieval RI et al. Lessons learned combining N-of-1 trials to assess fibromyalgia therapies. Journal of Rheumatology 2006; 33(10):2069-2077.
10.	Nikles CJ, Mitchell GK, Del Mar CB, Clavarino A, McNairn N. An n-of-1 trial service in clinical practice: testing the effectiveness of stimulants for attention-deficit/hyperactivity disorder. Pediatrics 2006; 117(6):2040-2046.
11.	Nathan PC, Tomlinson G, Dupuis LL, Greenberg ML, Ota S, Bartels U et al. A pilot study of ondansetron plus metopimazine vs. ondansetron monotherapy in children receiving highly emetogenic chemotherapy: a Bayesian randomized serial N-of-1 trials design. Supportive Care in Cancer 2006; 14(3):268-276.
12.	Pereira JA, Holbrook AM, Dolovich L, Goldsmith C, Thabane L, Douketis JD et al. Are brand-name and generic warfarin interchangeable? Multiple n-of-1 randomized, crossover trials. Annals of Pharmacotherapy 2005; 39(7-8):1188-1193.
13.	Woodfield R, Goodyear-Smith F, Arroll B. N-of-1 trials of quinine efficacy in skeletal muscle cramps of the leg. British Journal of General Practice 2005; 55(512):181-185.
14.	Wegman AC, van der Windt DA, Bongers M, Twisk JW, Stalman WA, de Vries TP. Efficacy of temazepam in frequent users: a series of N-of-1 trials. Family Practice 2005; 22(2):152-159.
15.	Nikles CJ, Yelland M, Glasziou PP, Del MC. Do individualized medication effectiveness tests (n-of-1 trials) change clinical decisions about which drugs to use for osteoarthritis and chronic pain?. [Review] [19 refs]. American Journal of Therapeutics 2005; 12(1):92-97.
16.	Smith BJ, Appleton SL, Veale AJ, McElroy HJ, Veljkovic D, Saccoia L. Eformoterol n-of-1 trials in chronic obstructive pulmonary disease poorly reversible to salbutamol. Chronic Respiratory Disease 2004; 1(2):63-69.

17.	Haas DC, Sheehe PR. Dextroamphetamine pilot crossover trials and n of 1 trials in patients with chronic tension-type and migraine headache. Headache 2004; 44(10):1029-1037.
18.	Mandelcorn J, Cullen NK, Bayley MT. A preliminary study of the efficacy of ondansetron in the treatment of ataxia, poor balance and incoordination from brain injury. Brain Injury 2004; 18(10):1025-1039.
19.	Pope JE, Prashker M, Anderson J. The efficacy and cost effectiveness of N of 1 studies with diclofenac compared to standard treatment with nonsteroidal antiinflammatory drugs in osteoarthritis. Journal of Rheumatology 2004; 31(1):140-149.
20.	Wegman AC, van der Windt DA, de HM, Deville WL, Fo CT, de Vries TP. Switching from NSAIDs to paracetamol: a series of n of 1 trials for individual patients with osteoarthritis. Annals of the Rheumatic Diseases 2003; 62(12):1156-1161.
21.	Reitberg DP, Del RE, Weiss SL, Rebell G, Zaias N. Single-patient drug trial methodology for allergic rhinitis. Annals of Pharmacotherapy 2002; 36(9):1366-1374.
22.	Linday LA, Tsiouris JA, Cohen IL, Shindledecker R, DeCresce R. Famotidine treatment of children with autistic spectrum disorders: pilot research using single subject research design. Journal of Neural Transmission 2001; 108(5):593-611.
23.	Duggan CM, Mitchell G, Nikles CJ, Glasziou PP, Del Mar CB, Clavarino A. Managing ADHD in general practice. N of 1 trials can help! Australian Family Physician 2000; 29(12):1205-1209.
24.	Nikles CJ, Glasziou PP, Del Mar CB, Duggan CM, Clavarino A, Yelland MJ. Preliminary experiences with a single-patient trials service in general practice. Medical Journal of Australia 2000; 173(2):100-103.
25.	Mahon JL, Laupacis A, Hodder RV, McKim DA, Paterson NA, Wood TE et al. Theophylline for irreversible chronic airflow limitation: a randomized study comparing n of 1 trials to standard practice. Chest 1999; 115(1):38-48.
26.	Bollert FG, Paton JY, Marshall TG, Calvert J, Greening AP, Innes JA. Recombinant DNase in cystic fibrosis: a protocol for targeted introduction through n-of-1 trials. Scottish Cystic Fibrosis Group. European Respiratory Journal 1999; 13(1):107-113.
27.	Kent MA, Camfield CS, Camfield PR. Double-blind methylphenidate trials: practical, useful, and highly endorsed by families. Archives of Pediatrics & Adolescent Medicine 1999; 153(12):1292-1296.
28.	Webb S, Tansey P, Brown H, Jackson A, Bilton D. Placebo-controlled n-of-1 trials in cystic fibrosis. European Respiratory Journal 1999; 14(4):993.
29.	Haines DR, Gaines SP. N of 1 randomised controlled trials of oral ketamine in patients with chronic pain. Pain 1999; 83(2):283-287.
30.	Sheather-Reid RB, Cohen M. Efficacy of analgesics in chronic pain: a series of N-of-1 studies. Journal of Pain & Symptom Management 1998; 15(4):244-252.
31.	Camfield P, Gordon K, Dooley J, Camfield C. Melatonin appears ineffective in children with intellectual deficits and fragmented sleep: six "N of 1" trials. Journal of Child Neurology 1996; 11(4):341-343.
32.	Mahon J, Laupacis A, Donner A, Wood T. Randomised study of n of 1 trials versus standard practice.[Erratum appears in BMJ 1996 Jun 1;312(7043):1392]. BMJ 1996; 312(7038):1069-1074.
33.	Maier W, Benkert O. Treatment of chronic depression with sulpiride: evidence of efficacy in placebo-controlled single case studies. Psychopharmacology 1994; 115(4):495-501.
34.	McQuay HJ, Carroll D, Jadad AR, Glynn CJ, Jack T, Moore RA et al. Dextromethorphan for the treatment of neuropathic pain: a double-blind randomised controlled crossover trial with integral n-of-1 design. Pain 1994; 59(1):127-133.
35.	March L, Irwig L, Schwarz J, Simpson J, Chock C, Brooks P. n of 1 trials comparing a non-steroidal anti-inflammatory drug with paracetamol in osteoarthritis. BMJ 1994; 309(6961):1041-1045.
36.	Denburg SD, Carbotte RM, Denburg JA. Corticosteroids and neuropsychological functioning in

	patients with systemic lupus erythematosus. Arthritis & Rheumatism 1994; 37(9):1311-1320.
37.	Privitera MD, Treiman DM, Pledger GW, Sahlroot JT, Handforth A, Linde MS et al. Dezinamide for partial seizures: results of an n-of-1 design trial. Neurology 1994; 44(8):1453-1458.
38.	Langer JC, Winthrop AL, Issenman RM. The single-subject randomized trial. A useful clinical tool for assessing therapeutic efficacy in pediatric practice. Clinical Pediatrics 1993; 32(11):654-657.
39.	Molloy DW, Guyatt GH, Standish T, Willan A, McIlroy W, D'Souza J et al. Effect of a new nootropic agent, CGS 5649B, on cognition, function, and behavior in dementia. Journal of General Internal Medicine 1993; 8(8):444-447.
40.	Johannessen T, Petersen H, Kristensen P, Fosstvedt D, Kleveland PM, Dybdahl J et al. Cimetidine on-demand in dyspepsia. Experience with randomized controlled single-subject trials. Scandinavian Journal of Gastroenterology 1992; 27(3):189-195.
41.	Johannessen T, Kristensen P, Petersen H, Fosstvedt D, Loge I, Kleveland PM et al. The symptomatic effect of 1-day treatment periods with cimetidine in dyspepsia. Combined results from randomized, controlled, single-subject trials. Scandinavian Journal of Gastroenterology 1991; 26(9):974-980.
42.	Patel A, Jaeschke R, Guyatt GH, Keller JL, Newhouse MT. Clinical usefulness of n-of-1 randomized controlled trials in patients with nonreversible chronic airflow limitation. American Review of Respiratory Disease 1991; 144(4):962-964.
43.	Larsen S, Farup P, Flaten O, Osnes M. The multi-crossover model for classifying patients as responders to a given treatment. Scandinavian Journal of Gastroenterology 1991; 26(7):763-770.
44.	Jaeschke R, Adachi J, Guyatt G, Keller J, Wong B. Clinical usefulness of amitriptyline in fibromyalgia: the results of 23 N-of-1 randomized controlled trials. Journal of Rheumatology 1991; 18(3):447-451.
45.	Hinderer SR. The supraspinal anxiolytic effect of baclofen for spasticity reduction. American Journal of Physical Medicine & Rehabilitation 1990; 69(5):254-258.
46.	Lashner BA, Hanauer SB, Silverstein MD. Testing nicotine gum for ulcerative colitis patients. Experience with single-patient trials. Digestive Diseases & Sciences 1990; 35(7):827-832.
47.	McBride MC. An individual double-blind crossover trial for assessing methylphenidate response in children with attention deficit disorder. Journal of Pediatrics 1988; 113(1:Pt 1):t-45.
48.	Menard J, Serrurier D, Bautier P, Plouin PF, Corvol P. Crossover design to test antihypertensive drugs with self-recorded blood pressure. Hypertension 1988; 11(2):153-159.
49.	Ullmann RK, Sleator EK. Responders, nonresponders, and placebo responders among children with attention deficit disorder. Importance of a blinded placebo evaluation. Clinical Pediatrics 1986; 25(12):594-599.
50.	Wolfe B, Del RE, Weiss SL, Mendelson A, Elbaga TA, Huser FJ et al. Validation of a single-patient drug trial methodology for personalized management of gastroesophageal reflux disease. J Manag Care Pharm 2002; 8(6):459-468.
51.	Brookes ST, Biddle L, Paterson C, Woolhead G, Dieppe P. "Me's me and you's you": Exploring patients' perspectives of single patient (n-of-1) trials in the UK. Trials 2007; 8:10.
52.	Wallace AE, Kofoed LL. Statistical Analysis of Single Case Studies in the Clinical Setting: The Example of Methylphenidate Trials in Children with Attention-Deficit Hyperactivity Disorder. Journal of Child and Adolescent Psychopharmacology 1994; 4(3):141-150.
53.	Miyazaki S, Nonogi H, Goto Y, Sumiyoshi T, Haze K, Hiramori K. Comparison of the therapeutic efficacy of continuous and intermittent injection of isosorbide dinitrate: a randomized study on unstable angina. Internal Medicine 1995; 34(9):856-862.
54,	Parodi O, Maseri A, Simonetti I. Management of unstable angina at rest by verapamil. A double-blind cross-over study in coronary care unit. British Heart Journal 1979; 41(2):167-174.
55.	Parodi O, Simonetti I, Michelassi C, Carpeggiani C, Biagini A, L'Abbate A et al. Comparison of verapamil and propranolol therapy for angina pectoris at rest: a randomized, multiple-crossover, controlled trial in the coronary care unit. American Journal of Cardiology 1986; 57(11):899-906.

56.	Joy TR, Zou GY, Mahon JL. N-of-1 (single-patient) trials for statin-related myalgia. Annals of internal medicine 2014 Oct 7;161(7):531-2
57.	Lipka AF, Vrinten C, van Zwet EW, Schimmel KJM, Cornel MC, Kuijpers MR, et al. Ephedrine treatment for autoimmune myasthenia gravis. Neuromuscular disorders 2017; 27:259-265.
58.	McGarry ME, Illek B, Ly NP, Zlock L, Olshansky S, Moreno C, et al. In vivo and in vitro ivacaftor response in cystic fibrosis patients with residual CFTR function: n-of-1 studies. Pediatric pulmonology 2017;52(4):472-9.
59.	Mitchell GK, Hardy JR, Nikles CJ, Carmont SA, Senior HE, Schluter PJ, et al. The Effect of Methylphenidate on Fatigue in Advanced Cancer: An Aggregated N-of-1 Trial. Journal of pain and symptom management 2015 Sep;50(3):289-96.
60.	Nikles J, Mitchell GK, Hardy J, Agar M, Senior H, Carmont SA, et al. Testing pilocarpine drops for dry mouth in advanced cancer using n-of-1 trials: A feasibility study. Palliative Medicine 2015 Dec;29(10):967-74.
61.	Nikles J, Mitchell GK, Hardy J, Senior H, Carmont SA, Schluter PJ, et al. Single-patient multiple crossover studies to determine the effectiveness of paracetamol in relieving pain suffered by patients with advanced cancer taking regular opioids: A pilot study. 2016;30(8):800-2.
62.	Nikles J, Mitchell G, McKinlay L, Waugh MC, Epps A, Carmont SA, et al. A series of n-of-1 trials of stimulants in brain injured children. 2017;40(1):11-21.

Appendix Table 4: Risk of bias assessment

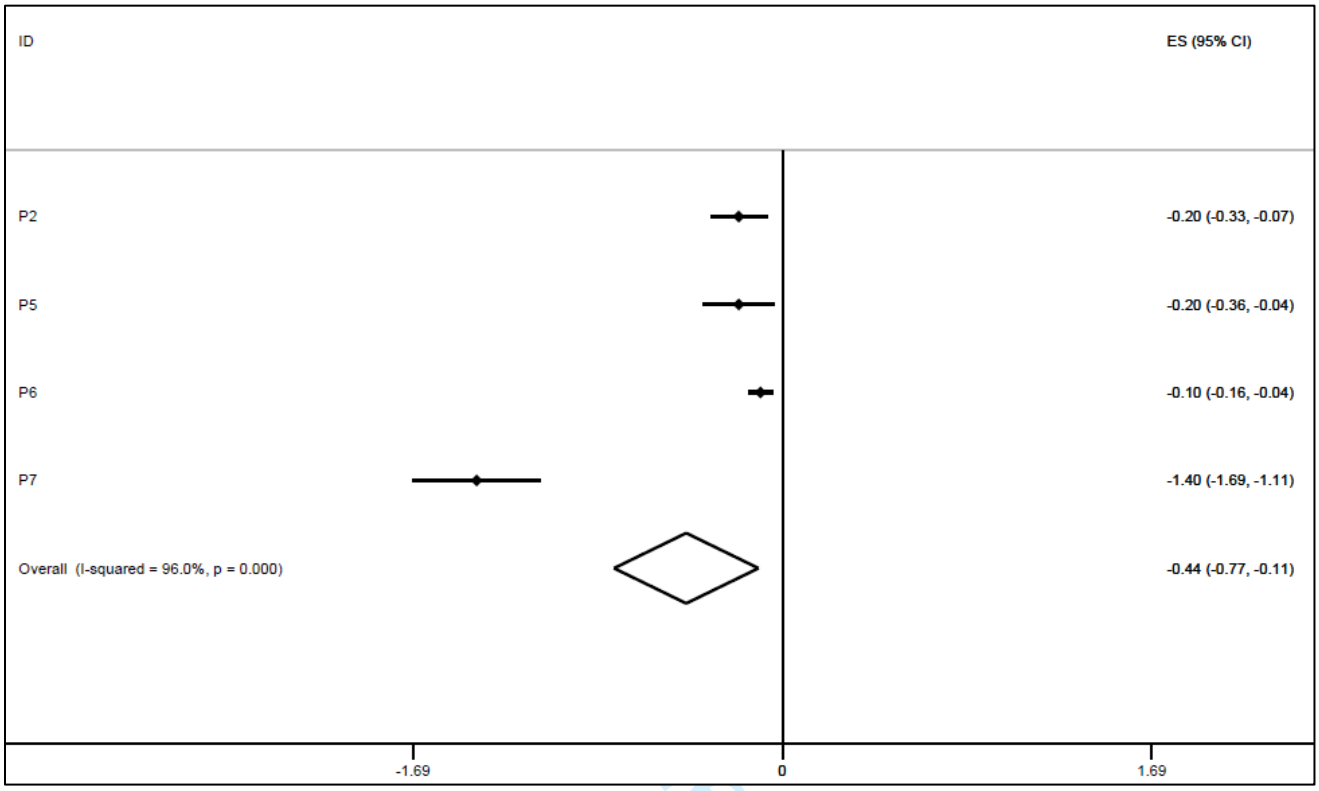
Author Yr	1. Randomization adequate?	2. Allocation concealed?	3. Patient blinded?	4. Outcome assessor blinded?	5. run-in period?	7. Wash-out?	8. Statistical methods appropriate?*	9. All randomized participants analyzed?	10. Incomplete outcome data
Nikles 2014	Low	Low	Low	Low	High	High	Low	High	Low
Tison 2013	Unclear	Low	Low	Low	High	Low	High	Low	Low
Rascol 2012	Unclear	Unclear	Low	Low	Low	Low	Low	High	Low
Emmanuel 2012	Unclear	Unclear	Low	Low	High	High	High	High	Low
Yelland 2009	Low	Low	Low	Low	High	High	Low	High	Low
Brookes 2007	Low	Low	Low	Low	High	High	unclear	High	Low
Nonoyama2007	Low	Low	Low	Low	High	High	unclear	High	Low
Huber 2007	Low	Low	Low	Low	High	Low	High	Low	Low
Yelland 2007	Low	Unclear	Low	Low	High	High	Low	High	Low
Zucker 2006	Low	Unclear	Low	Low	Low	High	Low	High	Low
Nikles 2006	Low	Unclear	Low	Low	High	Low	High	High	Low
Nathan 2006	Low	Low	Low	Low	High	High	High	High	Low
Pereira 1995	Unclear	Low	Low	Low	High	High	High	Low	Low
Woodfield 2005	Low	Low	Low	Low	Low	Low	Low	Low	Low
Wegman 2005	Low	Unclear	Low	Low	High	High	Low	High	Low
Nikles 2005	Low	Unclear	Low	Low	High	Low	High	High	Low
Smith 2004	Low	Low	Low	Low	Low	High	Low	High	Low
Haas 2004	Low	Low	Low	Low	High	High	Low	High	Low
Mandelcorn 2004	Low	Unclear	Low	Low	Low	High	High	Low	Low
Pope 2004	Unclear	High	Low	Low	High	High	Low	Low	Low
Wegman 2003	Low	Low	Low	Low	High	High	Low	High	Low

Wolfe 2002	Low	Low	Low	Low	High	Low	Low	High	Low
Reitberg 2002	Low	Low	Low	Low	Low	High	Low	Low	Low
Linday 2001	Unclear	Low	Low	Low	Low	High	High	High	Low
Duggan 2000	Unclear	Unclear	Low	Low	High	High	High	Low	Low
Nikles 2000	Unclear	Unclear	Low	Low	High	High	High	High	Low
Mahon 1999	Low	Unclear	Low	Low	High	Low	High	High	Low
Bollert 1999	Unclear	Unclear	Low	Low	High	High	High	High	Low
Kent 1999	Unclear	Unclear	Low	Low	High	High	High	Low	Low
Webb 1999	Unclear	Unclear	Low	Low	Low	Low	High	High	Low
Haines 1999	Unclear	Unclear	Low	Low	Low	Low	Low	High	Low
Sheather-Reid 1998	Unclear	Unclear	Low	Low	High	High	Low	High	Low
Camfield 1996	Unclear	Unclear	Low	Low	High	High	High	Low	Low
Mahon 1996	Low	Unclear	Low	Low	High	High	High	Low	Low
Miyazaki 1995	Unclear	High	High	High	High	High	High	High	Low
Maier 1994	Unclear	Unclear	Low	Low	Low	Low	Low	High	Low
McQuay 1994	Low	Low	Low	Low	High	High	High	High	Low
March 1994	Unclear	Unclear	Low	Low	High	High	Low	High	Low
Denburg 1994	Unclear	Unclear	Low	Low	Low	Low	High	High	Low
Privitera 1994	Unclear	Unclear	Low	Low	High	High	High	Low	Low
Wallace 1994	High	Unclear	Low	Low	High	High	High	Low	Low
Langer 1993	Low	Low	Low	Low	High	High	High	Low	Low
Molloy 1993	Unclear	Unclear	Low	Low	Low	Low	Low	High	Low
Johannessen 1992	Unclear	Unclear	Low	Low	High	Low	High	High	Low
Johannessen 1991	Unclear	Unclear	Low	Low	High	High	Low	Low	Low
Patel 1991	Unclear	Unclear	Low	Low	High	High	High	Low	Low
Larsen 1991	Unclear	Unclear	Low	Low	High	High	High	High	Low
Jaeschke 1991	Unclear	Unclear	Low	Low	low	High	High	High	low
Hinderer 1990	Unclear	Unclear	Low	Low	low	High	high	low	low
Lashner 1990	Unclear	Low	Low	Low	Unclear	High	high	low	low

McBride 1988	Low	Low	Low	Low	Unclear	High	high	low	High
Menard 1988	Low	Unclear	Low	Low	Low	low	low	low	High
Ullmann 1986	Low	Unclear	Low	Low	Unclear	High	low	low	High
Parodi 1986	Low	Unclear	Low	Low	low	Low	low	low	low
Parodi 1979	Unclear	Unclear	Low	Low	low	High	High	low	low
Joy 2014	Low	Unclear	Low	Low	High	Low	low	low	low
Lipka 2017	Low	Low	Low	Low	High	Low	High	low	low
Mitchell 2015	Low	Low	Low	Low	High	Low	High	low	low
Nikles 2015	Low	Low	Low	Low	High	Low	High	low	low
Nikles 2017	Low	Low	Low	Low	High	Low	low	High	low
Nikles 2016	Low	Unclear	Low	Low	High	High	High	low	High
McGarry 2017	Low	Low	Low	Low	Low	Low	High	High	High

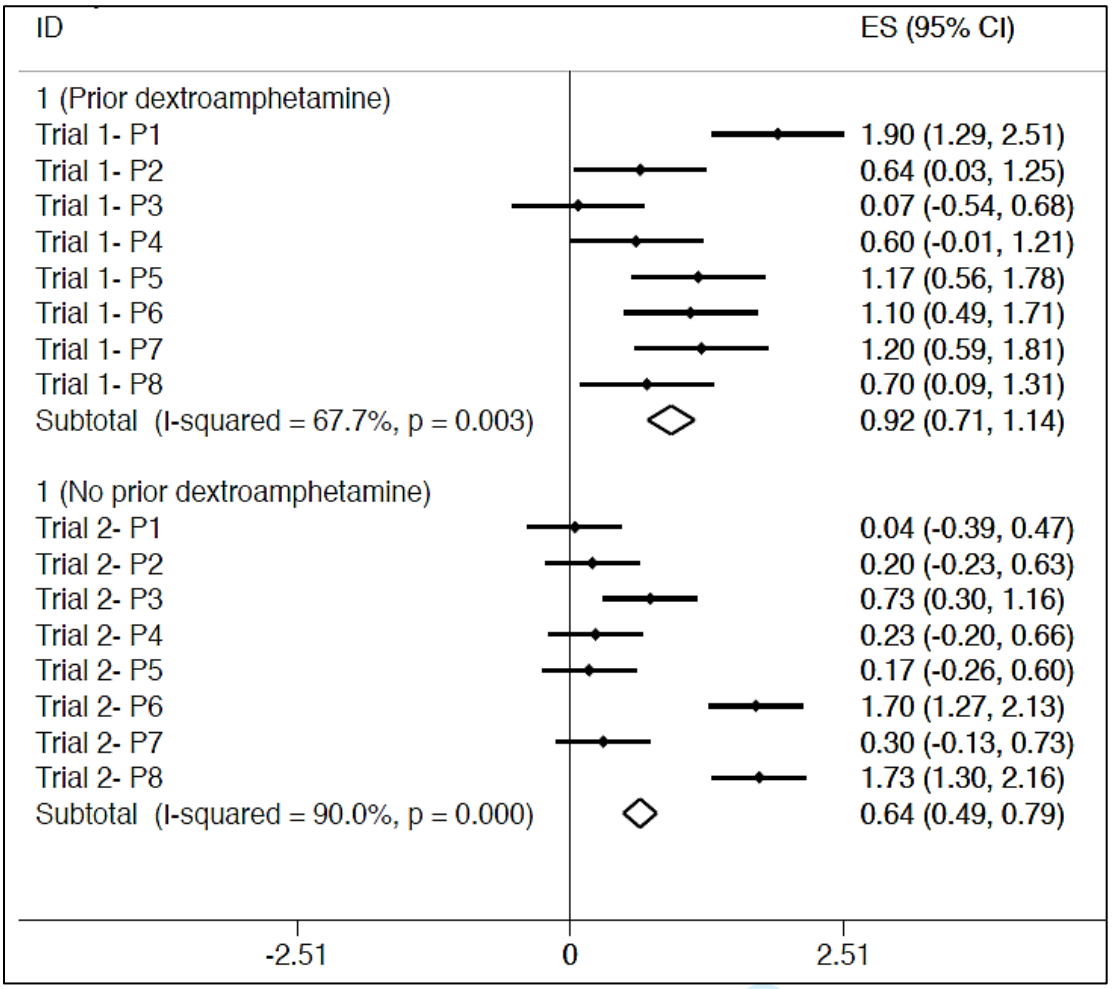
* Statistical methods used to account for carryover effect, period effects, and intra-subject correlation

Appendix Figure 1: Patients with chronic intestinal pseudo-obstruction treated with prucalopride or placebo for pain relief[1]



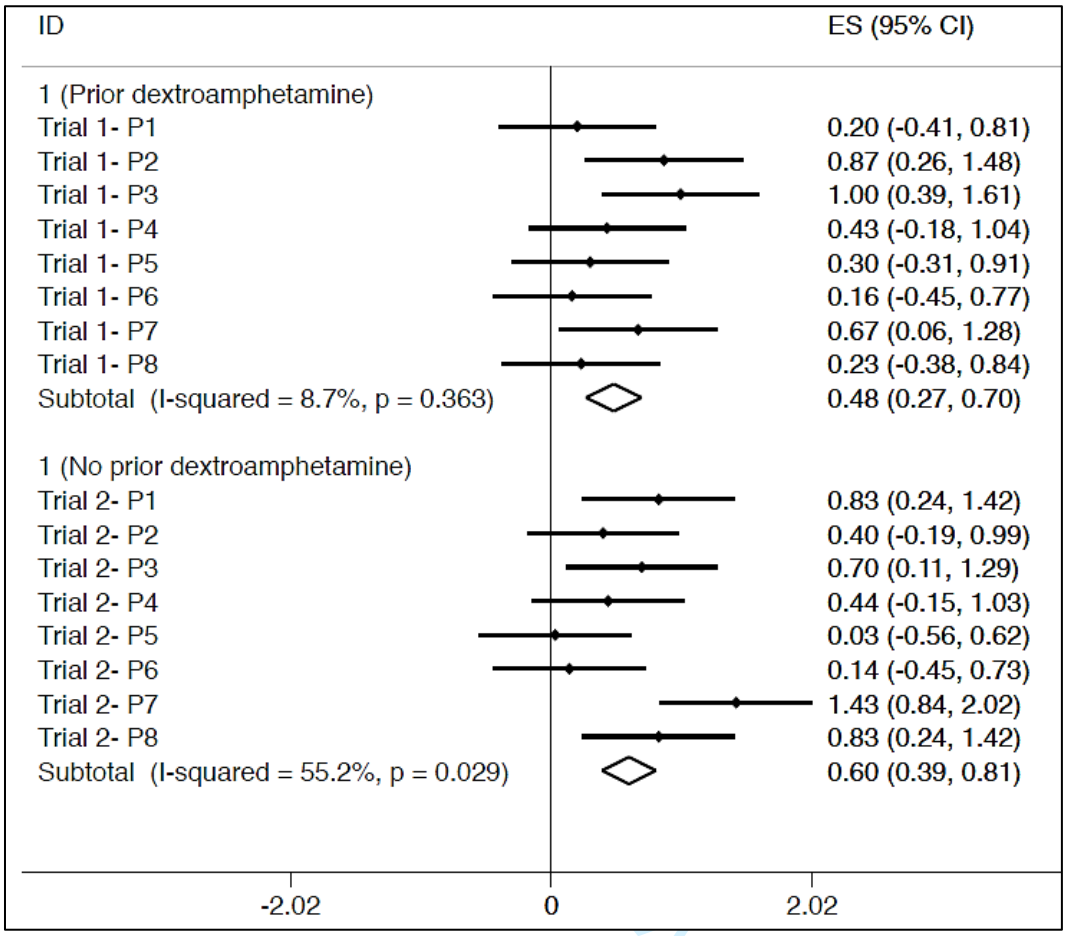
Appendix Figure 1 Legend:
Data from this figure was extracted from the study published by Emmanuel et al in 2011, which investigates the use of prucalopride or placebo for pain relief (among other outcomes) in patients with chronic intestinal pseudo-obstruction. The average treatment effect is -0.440 (-0.771 to -0.110).

Appendix Figure 2: Patients with chronic tension-type headaches treated with dextroamphetamine or control and effect on mean daily grade decrease in headache[2]



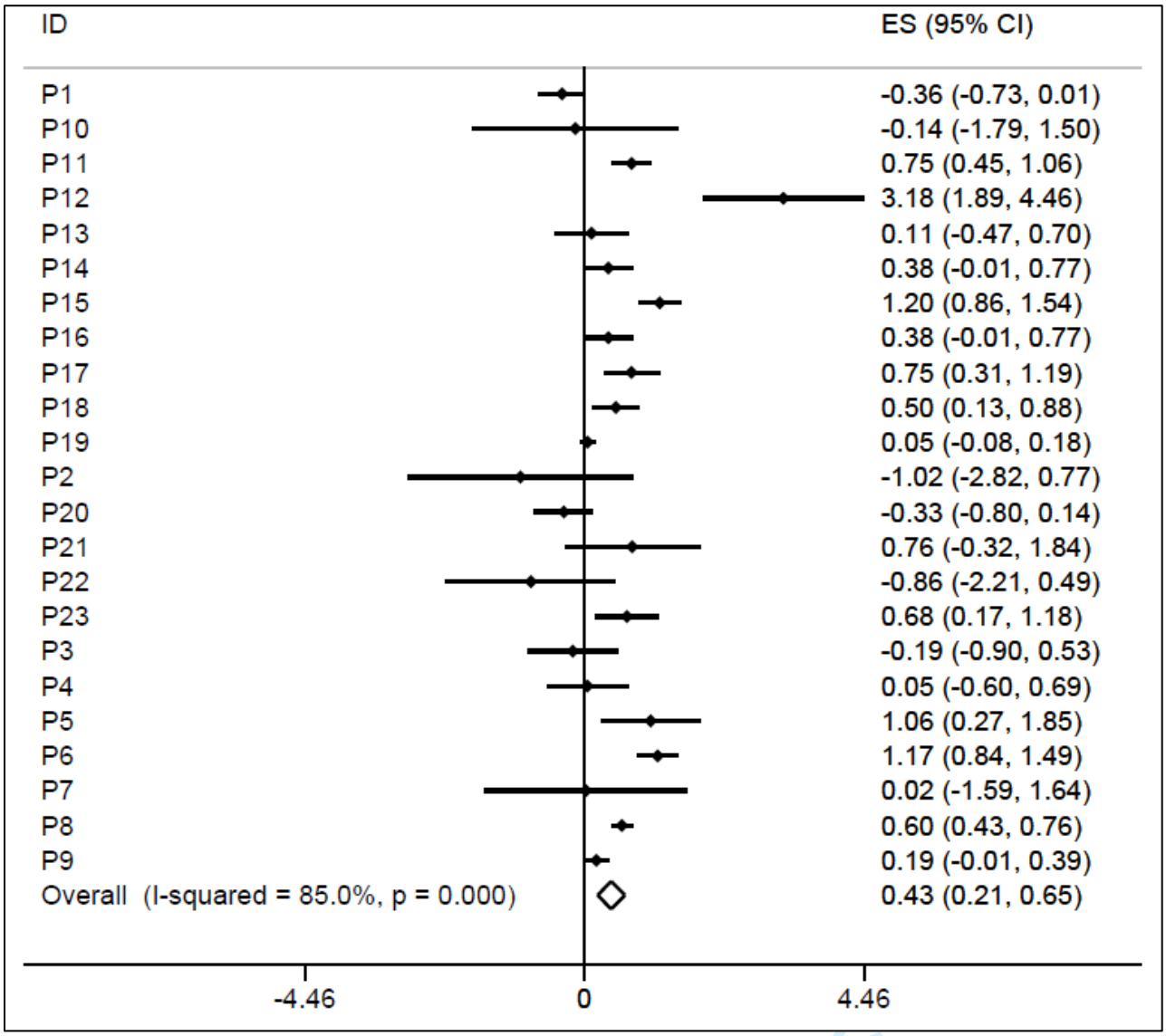
Appendix Figure 2 Legend: Data from this figure was extracted from the study published by Haas et al in 2004, which investigates the use of dextroamphetamine or control in patients with chronic-type for improvement on mean daily grade in headache.

Appendix Figure 3: Patients with migraine headaches treated with dextroamphetamine or control and effect on mean daily grade decrease in headache[2]



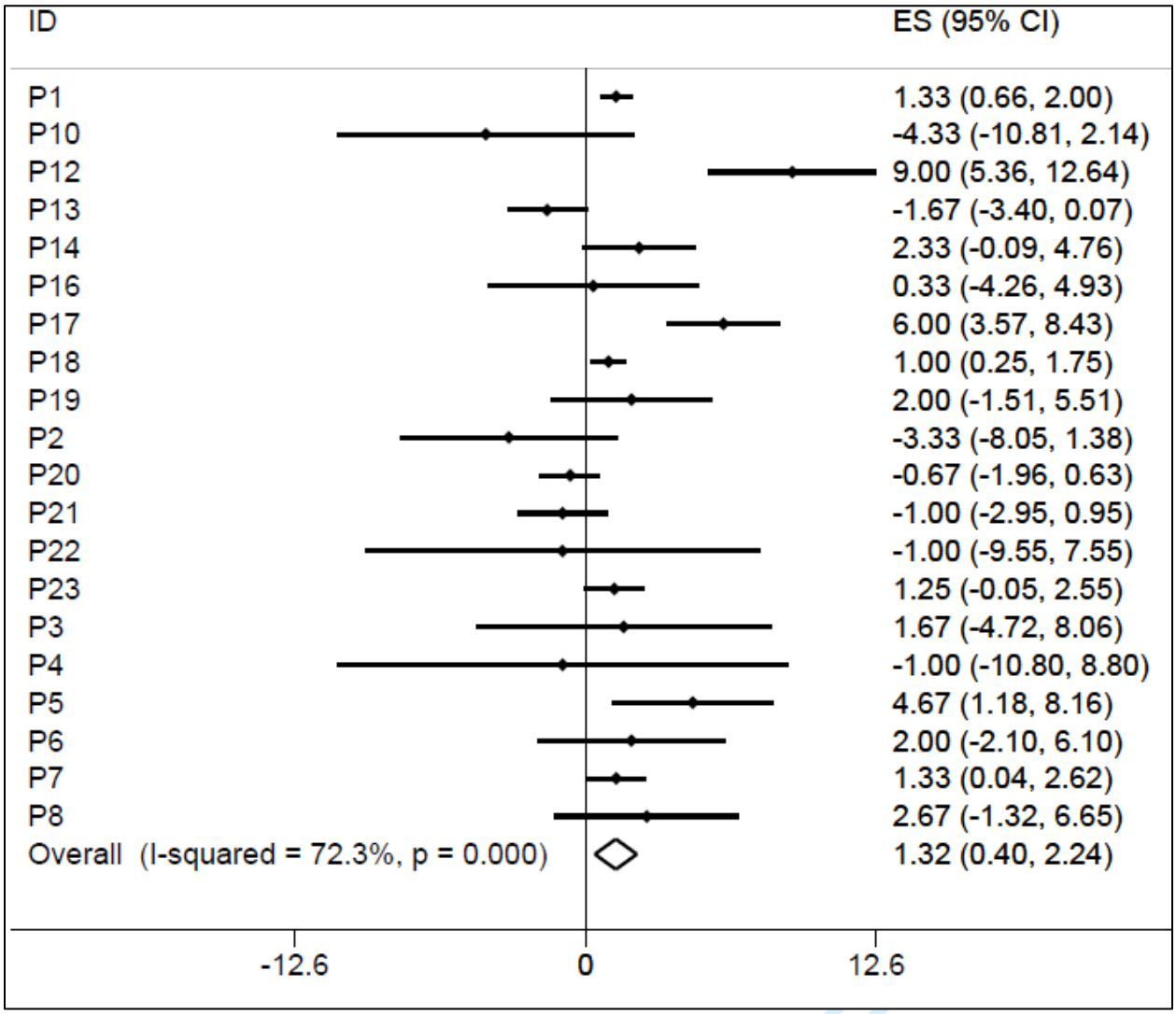
Appendix Figure 3 Legend: Data from this figure was extracted from the study published by Haas et al in 2004, which investigates the use of dextroamphetamine or control in patients with chronic-type and migraine headaches for improvement on mean daily grade in headache.

Appendix Figure 4: Patients with fibromyalgia treated with amitriptyline or placebo and its effect on a 7-point symptom scale[3]



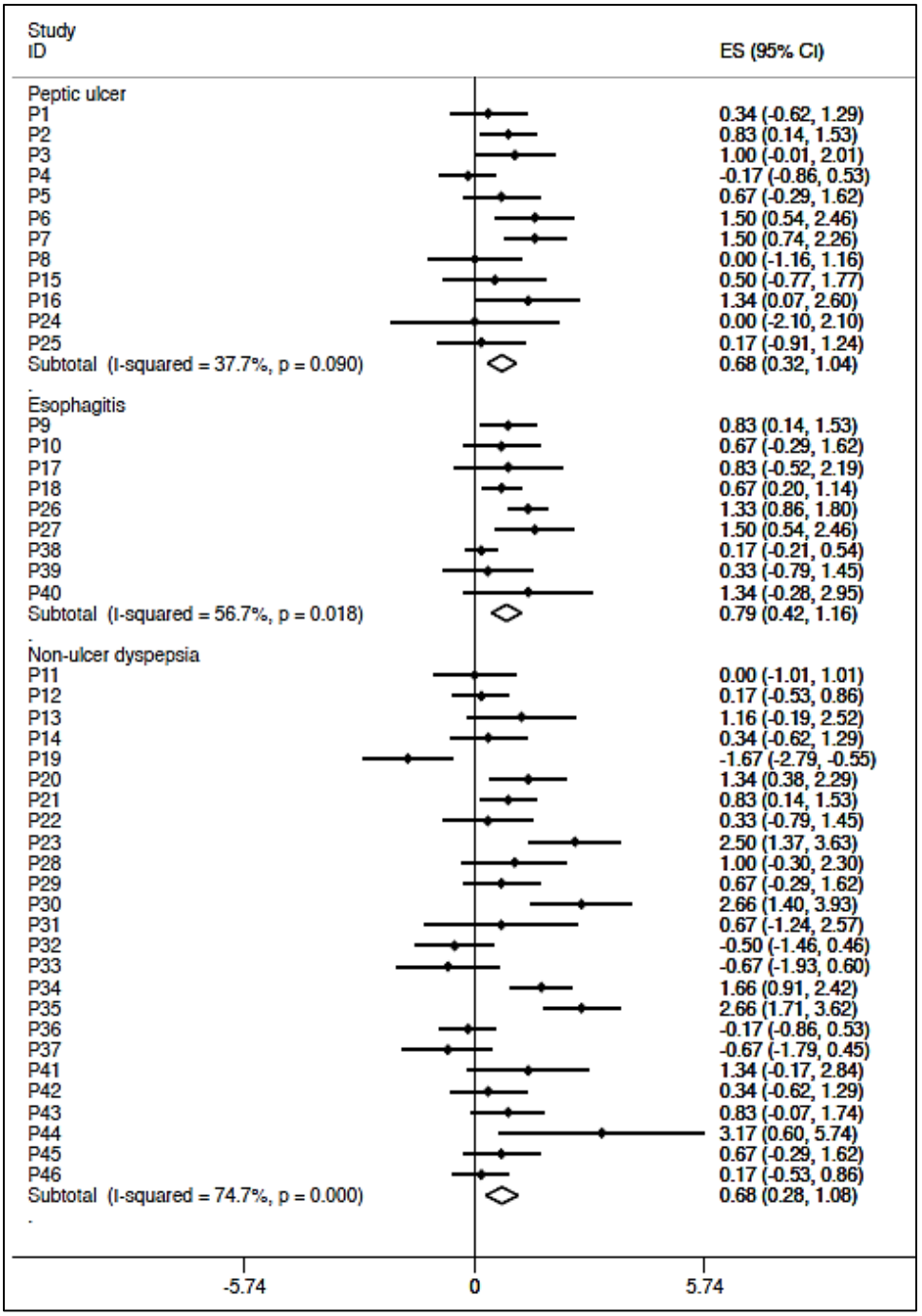
Appendix Figure 4 Legend: Data from this figure was extracted from the study published by Jaeschke et al in 1991, which investigates the effect of amitriptyline or placebo on a 7-point symptom scale in patients with fibromyalgia. The average treatment effect is 0.427 (0.210 to 0.645).

Appendix Figure 5: Patients with fibromyalgia treated with amitriptyline or placebo and its effect on tender point changes count[3]



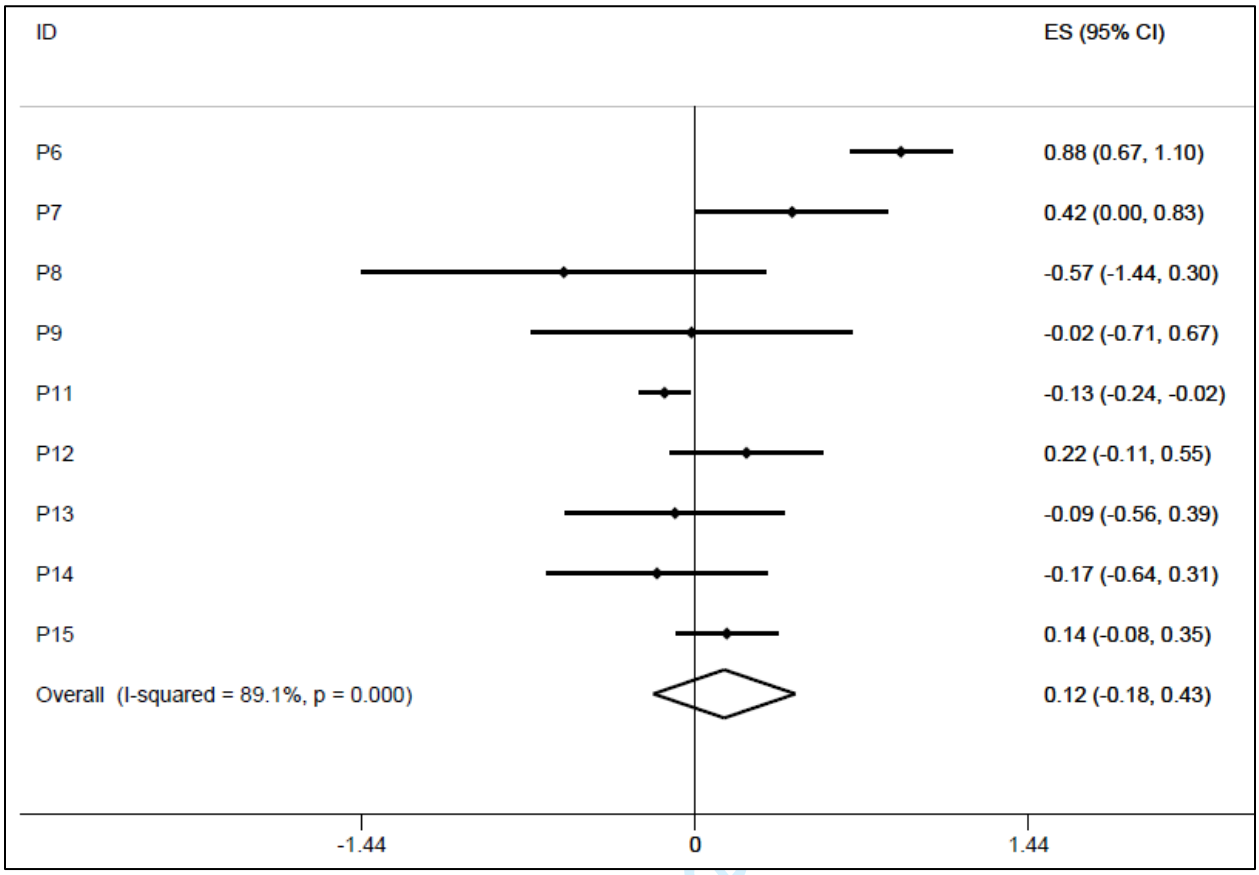
Appendix Figure 5 Legend: Data from this figure was extracted from the study published by Jaeschke et al in 1991, which investigates the effect of amitriptyline or placebo on tender point changes count in patients with fibromyalgia. The average treatment effect is 1.320 (0.404 to 2.236).

Appendix Figure 6: Patients with peptic ulcers, oesophagitis grade I, II, or III, or with reflux or ulcer-like symptom profiles were treated with cimetidine or placebo and its effect on a 6-point symptom scale[4]



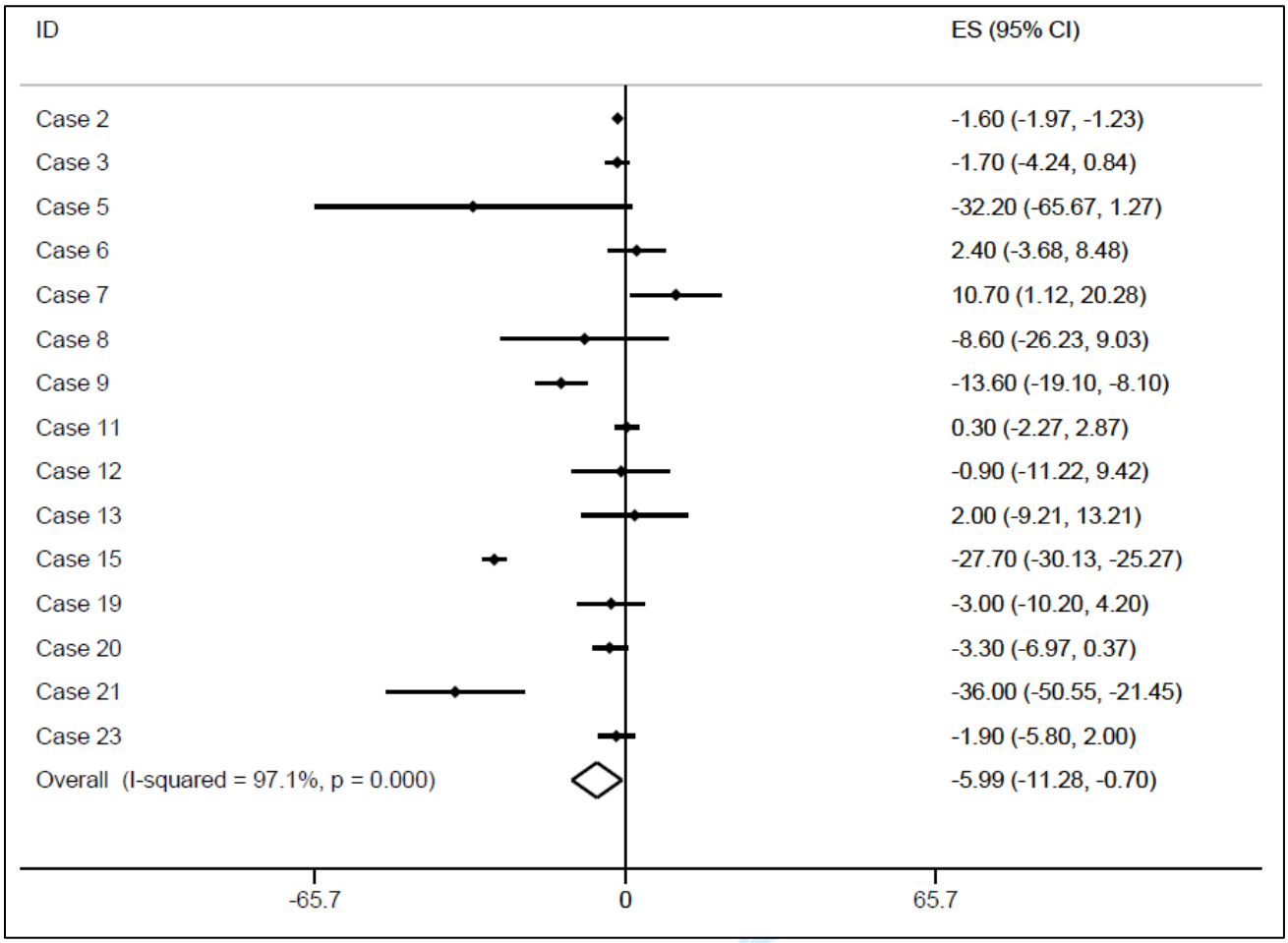
Appendix Figure 6 Legend: Data from this figure was extracted from the study published by Johannessen et al in 1992, which investigates the effect of cimetidine or placebo on a 6-point symptom scale in patients with peptic ulcers, oesophagitis grade I, II, or III, or with reflux or ulcer-like symptom profiles. The average treatment effect is 0.698 (0.466 to 0.931).

Appendix Figure 7: Patients with irreversible chronic airflow limitation treated with theophylline or placebo and its effect on dyspnea[5]



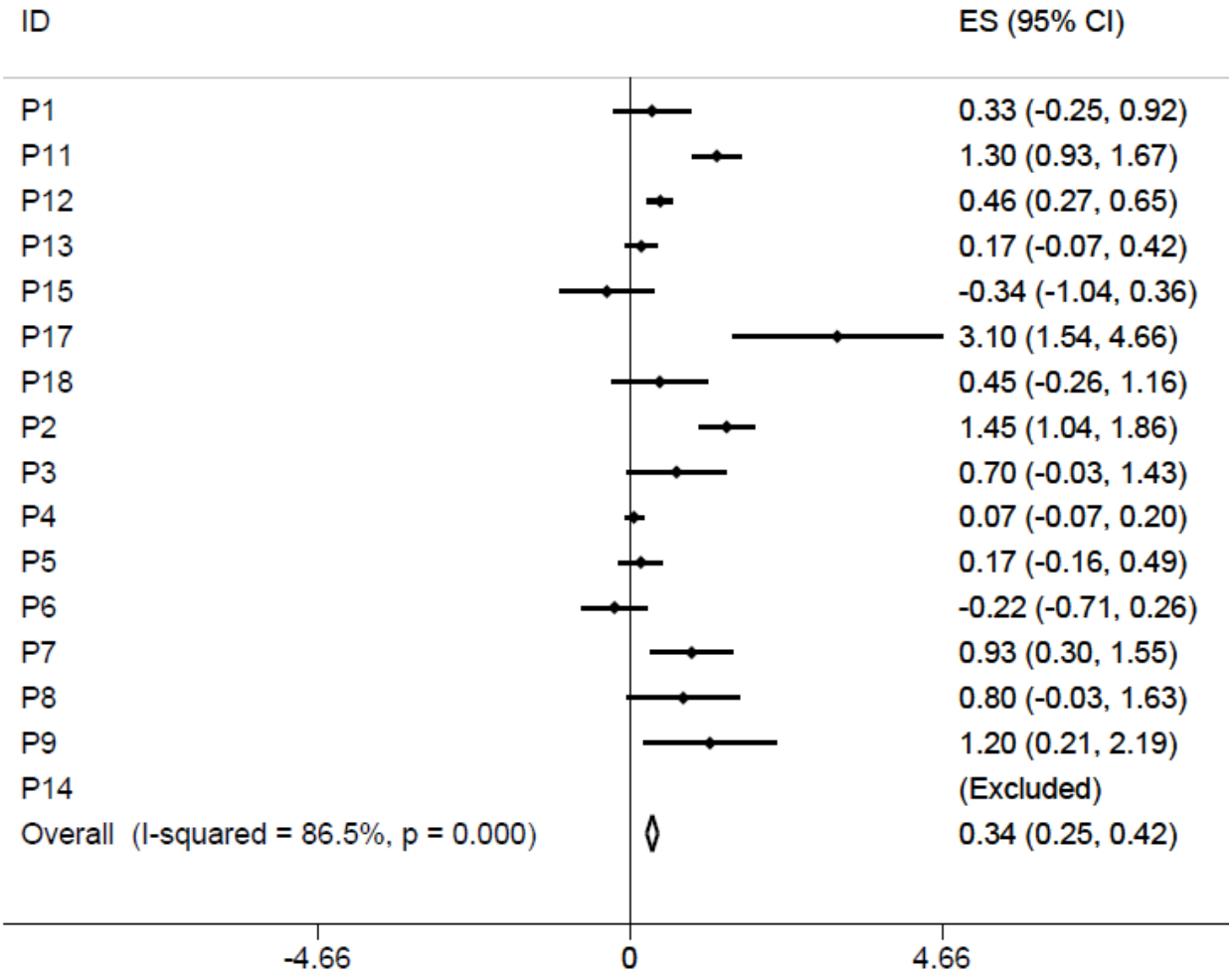
Appendix Figure 7 Legend: Data from this figure was extracted from the study published by Mahon et al in 1996, which investigates the effect of theophylline or placebo on dyspnea in patients with irreversible chronic airflow limitation. The average treatment effect is 0.125 (-0.181 to 0.430).

Appendix Figure 8: Patients with osteoarthritic pain treated with paracetmol and diclofenac and its effect on stiffness[6]



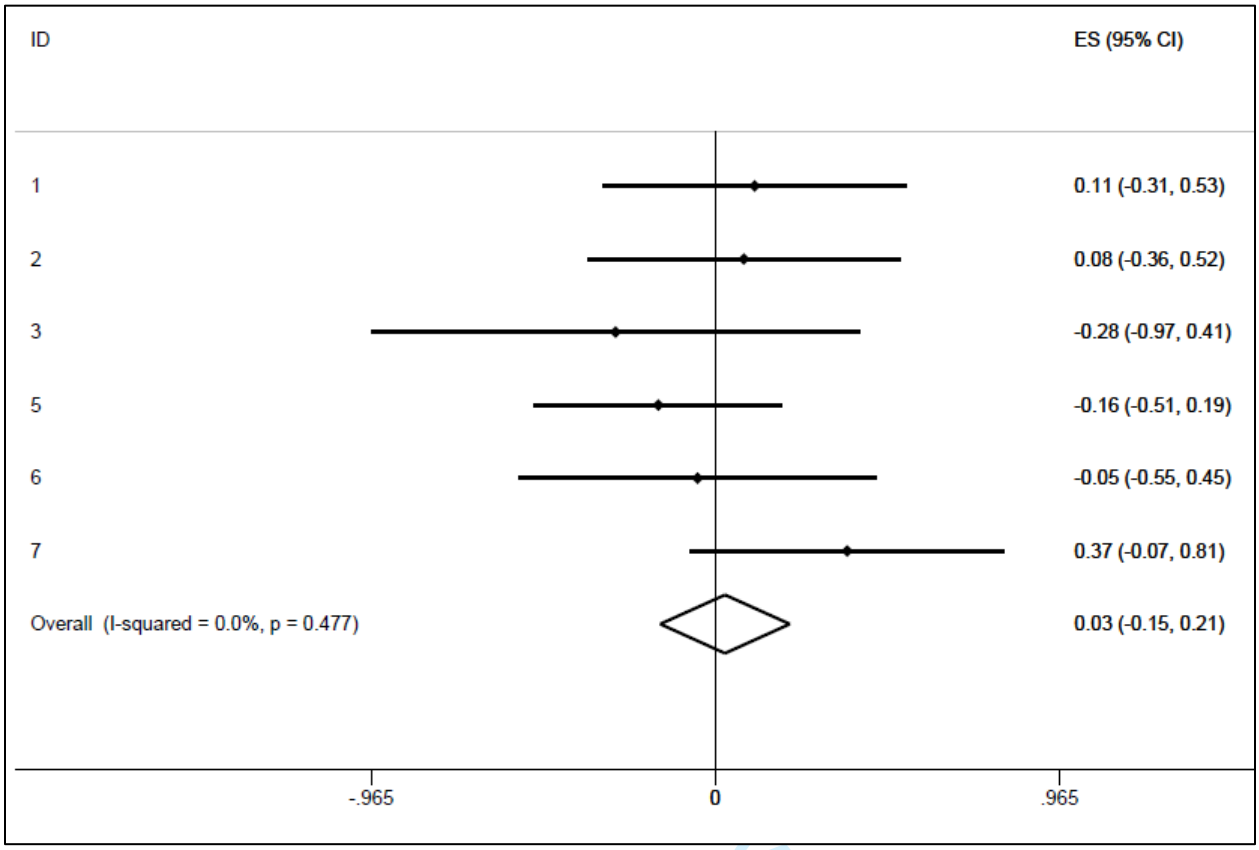
Appendix Figure 8 Legend: Data from this figure was extracted from the study published by March et al in 1994, which investigates the effect of paracetmol and diclofenac on stiffness in patients with osteoarthritic pain. The average treatment effect is mean difference in stiffness (mm).

Appendix Figure 9: Patients with nonreversible chronic airflow limitation treated with either ipratropium bromide, theophylline, salbutamol, or beclomethane (all compared to placebo) and its effect on a 4-item symptom questionnaire[7]



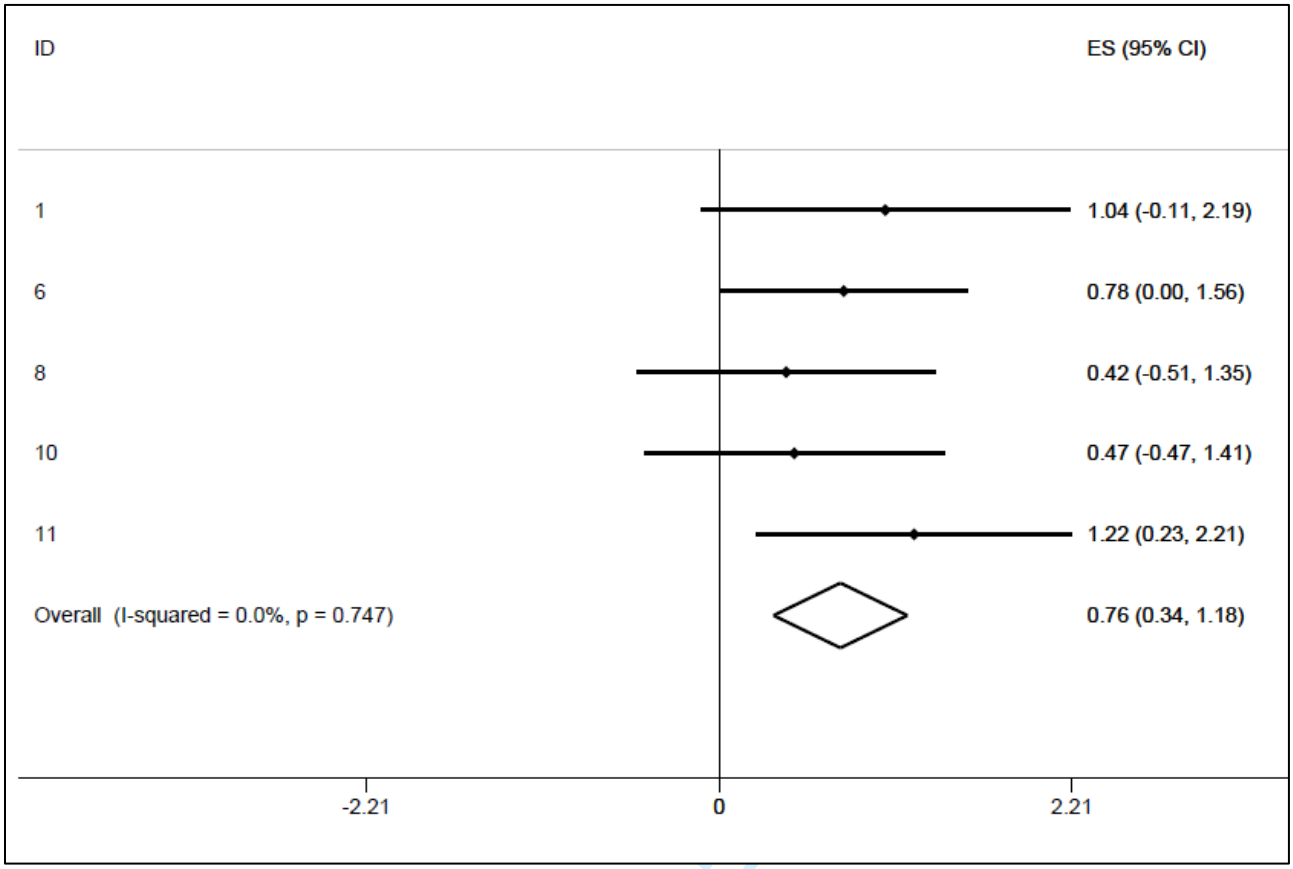
Appendix Figure 9 Legend: Data from this figure was extracted from the study published by Patel et al in 1991, which investigates the effect of ipratropium bromide, theophylline, salbutamol, or beclomethane (all compared to placebo) on a 4-item symptom questionnaire in patients with nonreversible chronic airflow limitation. The average treatment effect is 0.340 (0.253 to 0.422).

Appendix Figure 10: Patients previously taking warfarin for either atrial fibrillation or deep vein thrombosis treated with apo-warfarin and 20coumadin and its effect on international normalized ratio[8]



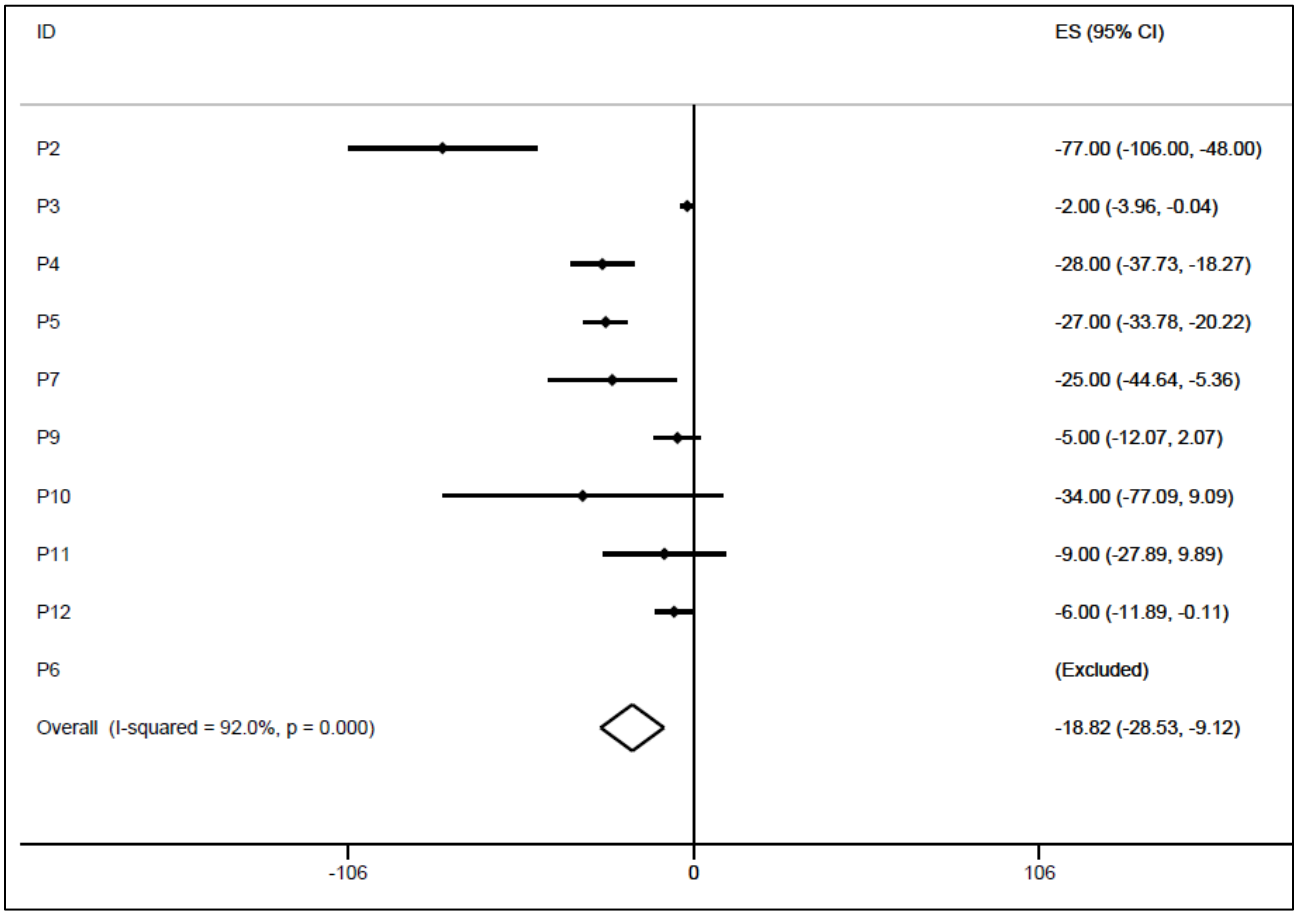
Appendix Figure 10 Legend: Data from this figure was extracted from the study published by Pereira et al in 1995, which investigates the effect of apo-warfarin and Coumadin on international normalized ratio in patients previously taking warfarin for either atrial fibrillation or deep vein thrombosis. The average treatment effect is 0.027 (-0.155 to 0.209).

Appendix Figure 11: Hospitalized children and adolescents with attention-deficit hyperactivity disorder treated with methylphenidate and placebo and its effect on Conners 15-item rating scale scores[9]



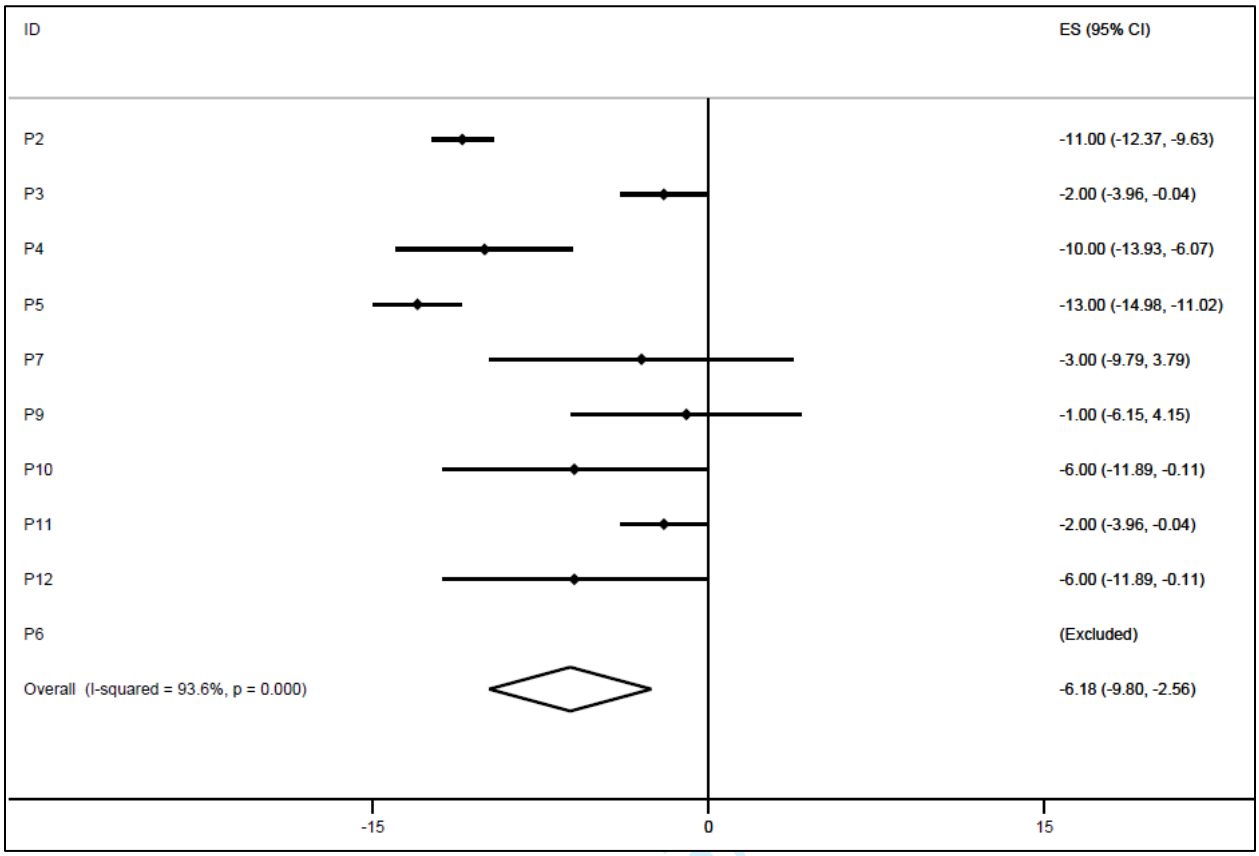
Appendix Figure 11 Legend: Data from this figure was extracted from the study published by Wallace et al in 1994, which investigates the effect of methylphenidate and placebo on Conners 15-item rating scale scores in hospitalized children and adolescents with attention-deficit hyperactivity disorder. The average treatment effect is 0.759 (0.341 to 1.178).

Appendix Figure 12: Patients already prescribed quinine treated with quinine sulphate and placebo, and its effect on changes in number of cramps[10]



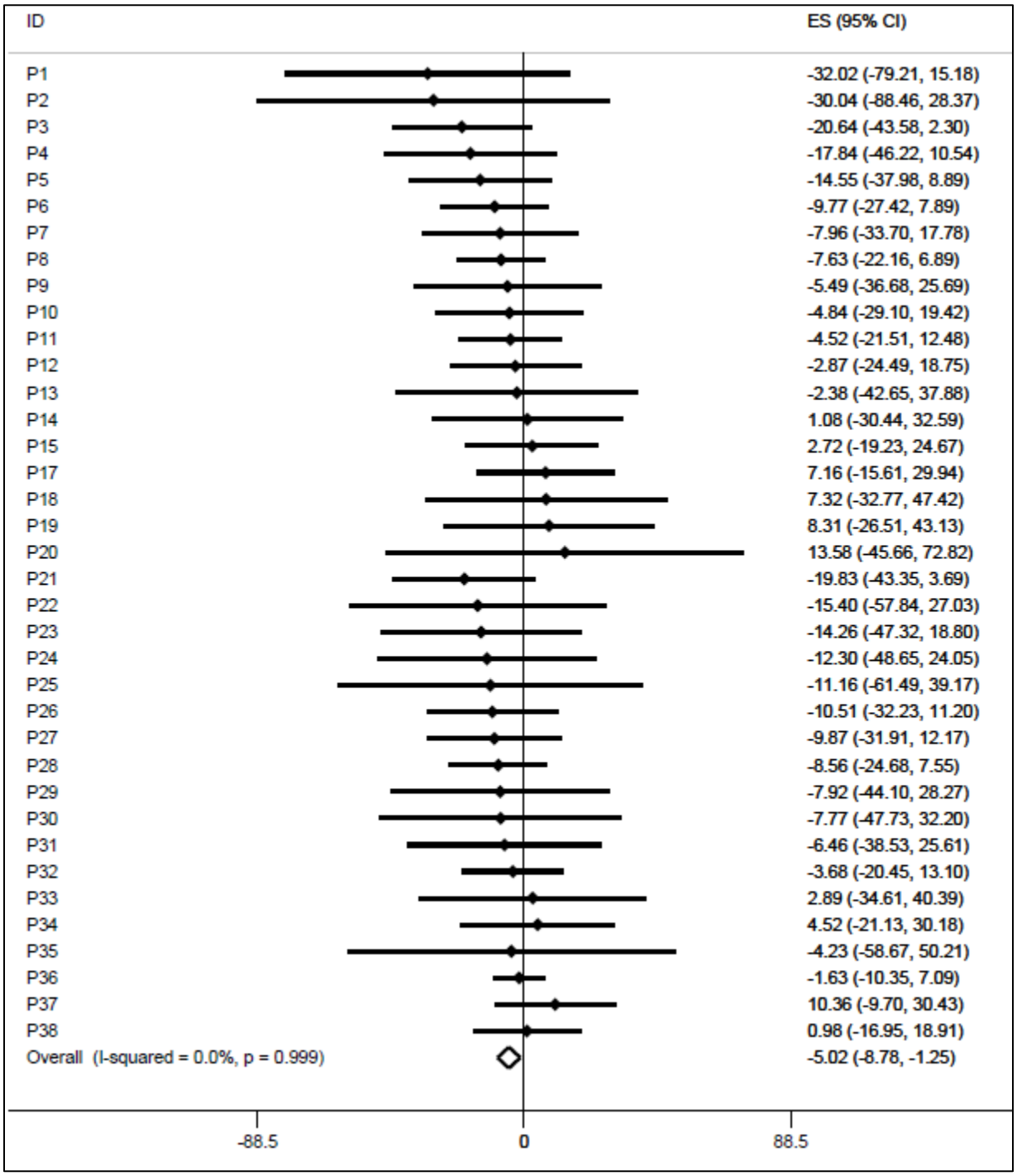
Appendix Figure 12 Legend: Data from this figure was extracted from the study published by Woodfield et al in 2005, which investigates the effect of quinine sulphate and placebo on changes in number of cramps in patients already prescribed quinine. The average treatment effect is -18.823 (-28.527 to -9.120).

Appendix Figure 13: Patients already prescribed quinine treated with quinine sulphate and placebo, and its effect on total days with cramps[10]



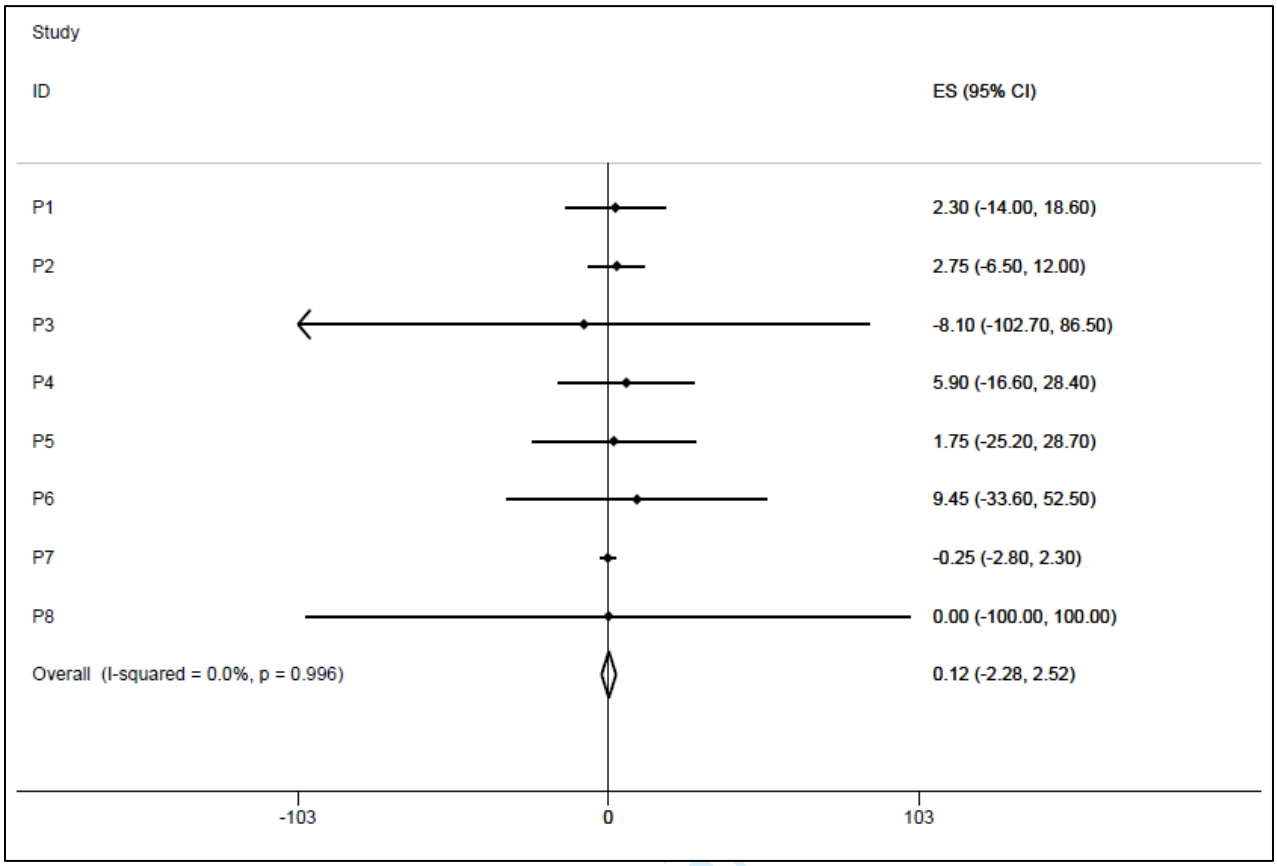
Appendix Figure 13 Legend: Data from this figure was extracted from the study published by Woodfield et al in 2005, which investigates the effect of quinine sulphate and placebo on total days with cramps in patients already prescribed quinine. The average treatment effect is -6.181 (-9.798 to -2.563).

Appendix Figure 14: Patients with fibromyalgia syndrome treated with amitriptyline and the combination amitriptyline and fluoxetine and its effect on the Fibromyalgia Impact Questionnaire[11]



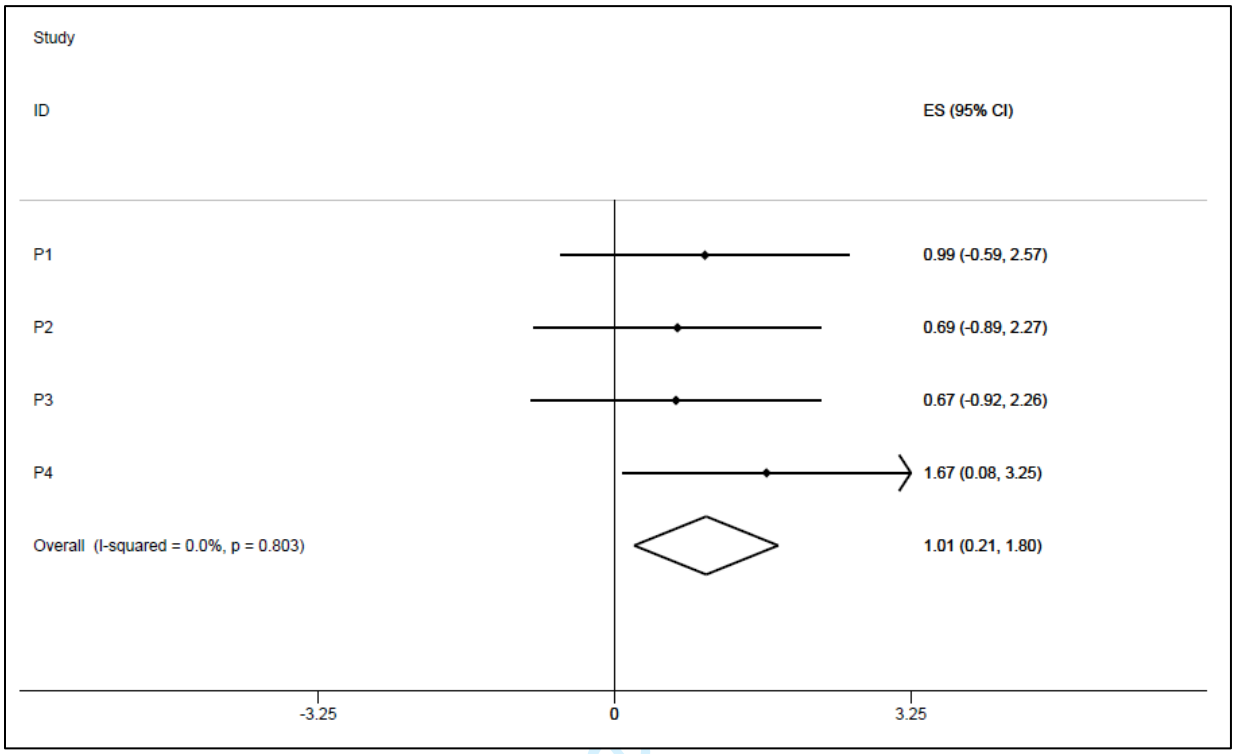
Appendix Figure 14 Legend: Data from this figure was extracted from the study published by Zucker et al in 2006, which investigates the effect of amitriptyline and the combination amitriptyline and fluoxetine on Fibromyalgia Impact Questionnaire in patients with fibromyalgia syndrome. The average treatment effect is -5.019 (-8.784 to -1.254).

Appendix 15: Patients with prior statin-related myalgia with or without mild elevation of creatine kinase levels treated with statin and placebo and its effects on VAS myalgia score[12]



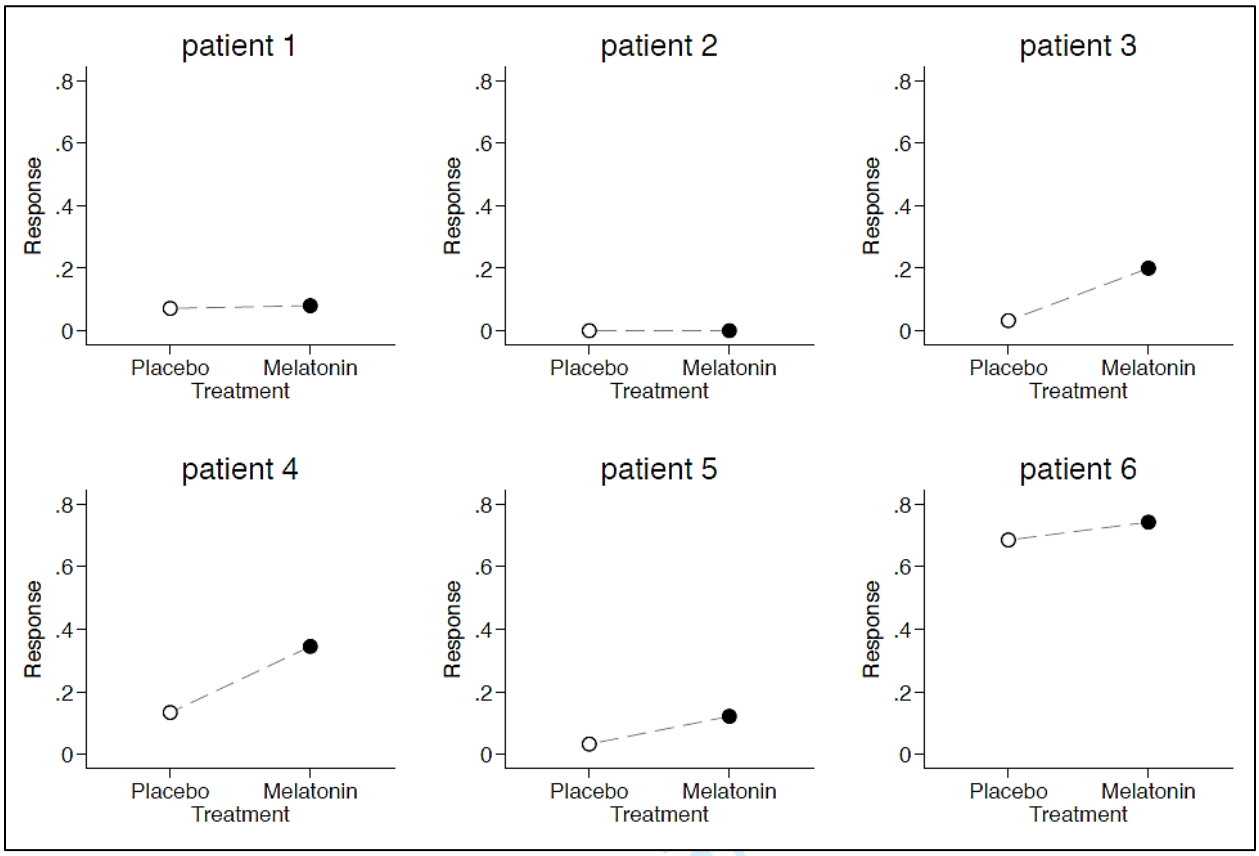
Appendix 15 Figure Legend: Data from this figure was extracted from the study published by Joy et al in 2014, which investigates the effect of statin versus placebo on VAS myalgia score in patients with hyperlipidemia. The average treatment effect is 0.12 (-2.28 to 2.52).

Appendix Figure 16: Patients with myasthenia gravis with acetylcholine receptor antibodies treated with ephinpherin and placebo and its effect on QMG score[13]



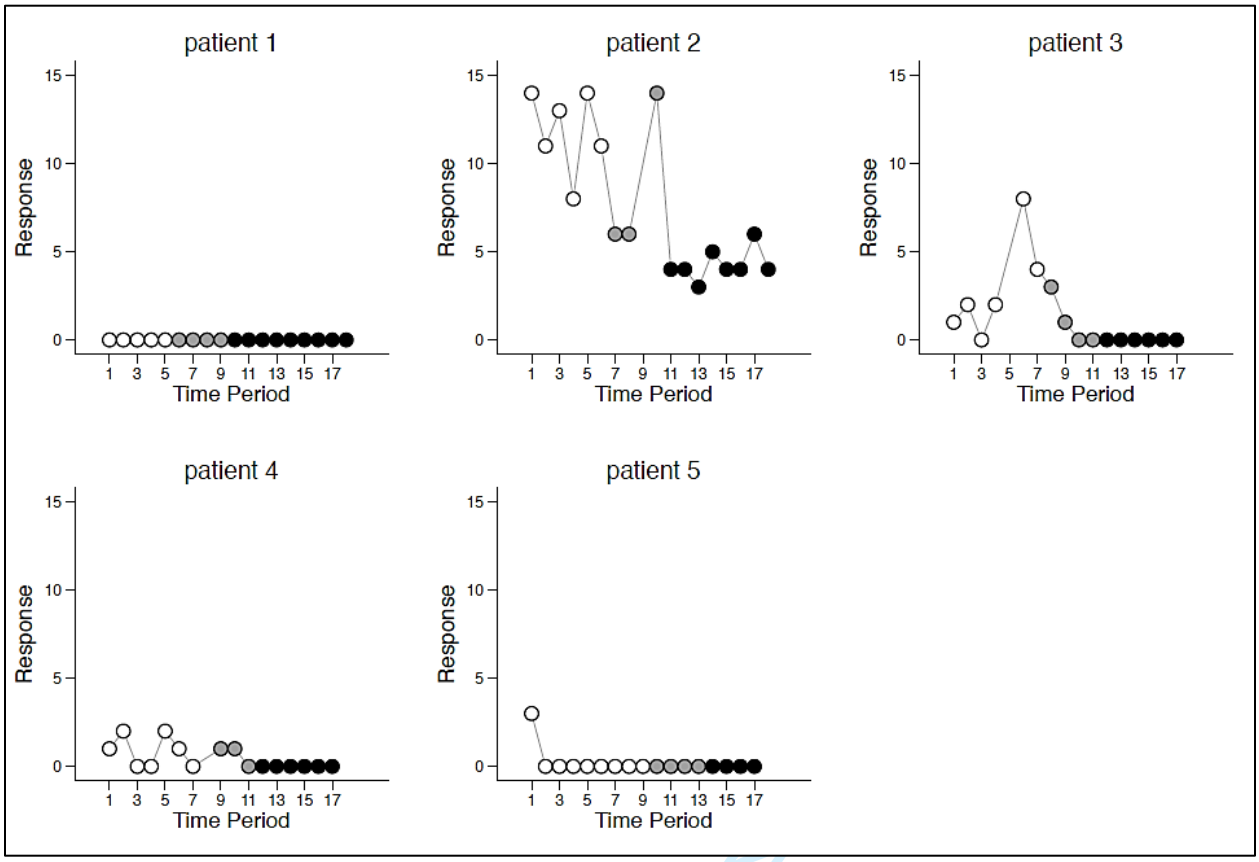
Appendix Figure 16 Legend: Data from this figure was extracted from the study published by Lipkin et al in 2017, which investigates the effect of with ephinpherin and placebo and its effect on QMG score in patients with autoimmune myasthenia gravia. The average treatment effect is 1.01 (0.21 to 1.80).

Appendix Figure 17: Children with mental retardation and fragmented sleep treated with melatonin and placebo and its effect on nights without awakening[14]



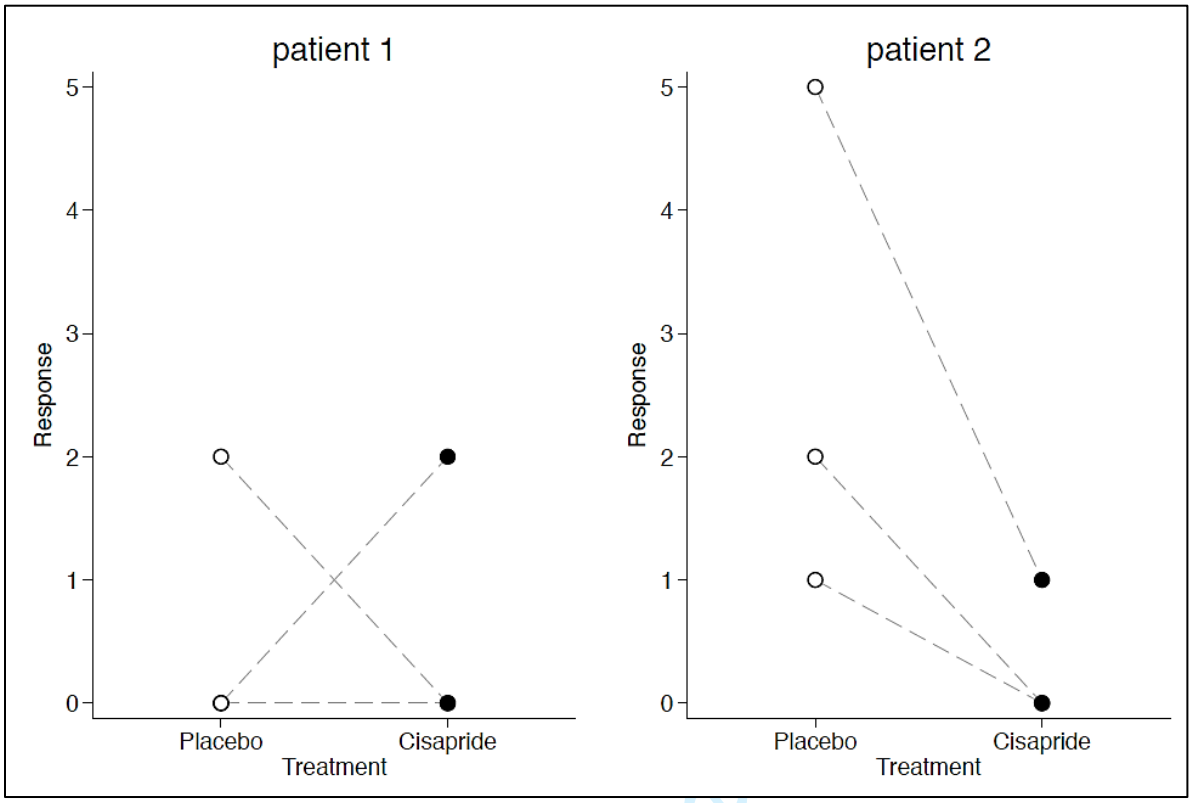
Appendix Figure 17 Legend: Data from this figure was extracted from the study published by Camfield et al in 1996, which investigates the effect of melatonin and placebo on nights without awakening in children with mental retardation and fragmented sleep. The average treatment effect is 0.84 (0.20 to 1.48). White circles indicate placebo; black circles indicate melatonin.

Appendix Figure 18: Patients with traumatic spinal cord lesions treated with baclofen and placebo and its effect on anxiety[15]



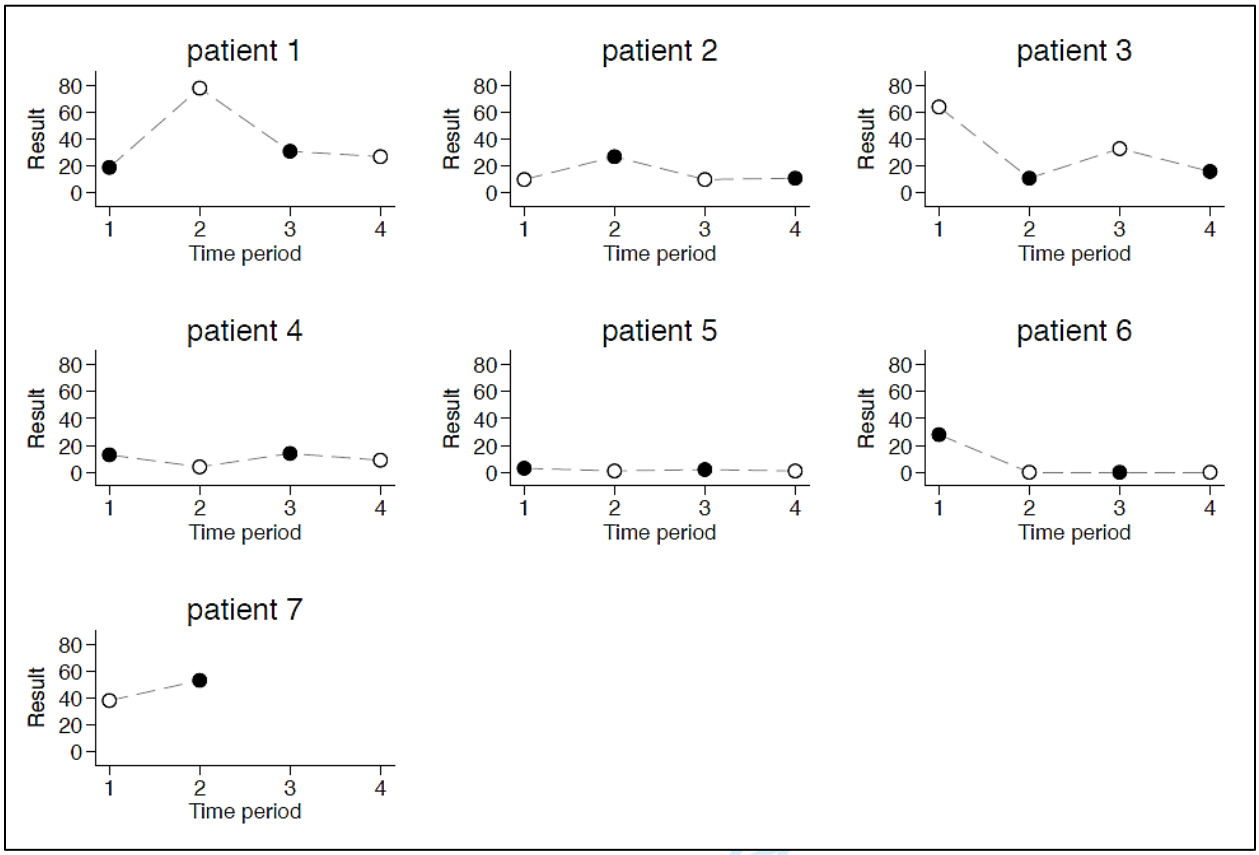
Appendix Figure 18 Legend: Data from this figure was extracted from the study published by Hinderer et al in 1990, which investigates the effect of baclofen and placebo on anxiety in patients with traumatic spinal cord lesions. The average treatment effect is -1.06 (-1.88 to -0.23). White circles indicate placebo; grey circles indicate a half dose (40 mg/day) of baclofen; black circles indicate a full dose (80 mg/day) of baclofen.

Appendix Figure 19: Children with gastroesophageal reflux treated with cisapride and placebo and its effect on emetic episodes per day[16]



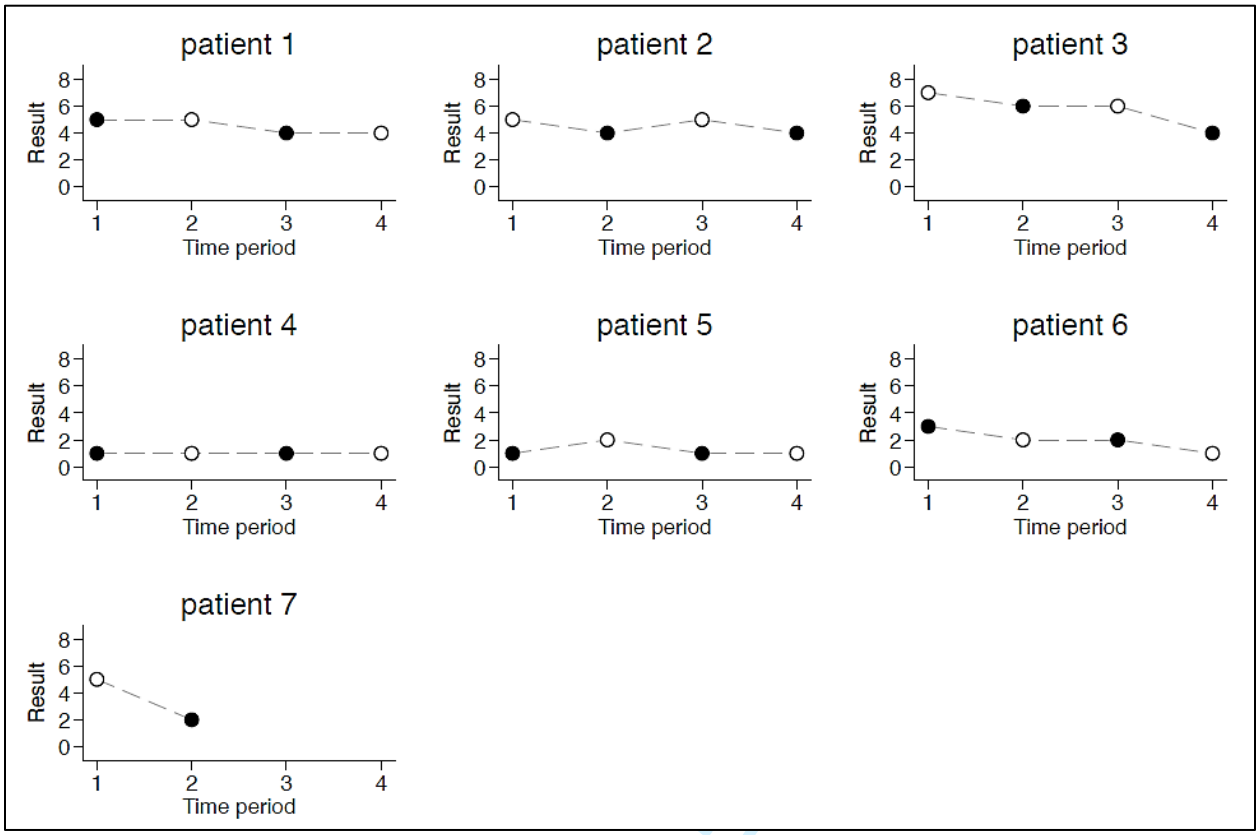
Appendix Figure 19 Legend: Data from this figure was extracted from the study published by Langer et al in 1993, which investigates the effect of cisapride and placebo on emetic episodes per day in children with gastroesophageal reflux. The average treatment effect is -1.20 (-2.49 to 0.09). White circles indicate placebo; black circles indicate cisapride.

Appendix Figure 20: Nonsmokers with ulcerative colitis treated with nicotine gum and placebo and its effect on abdominal pain[17]



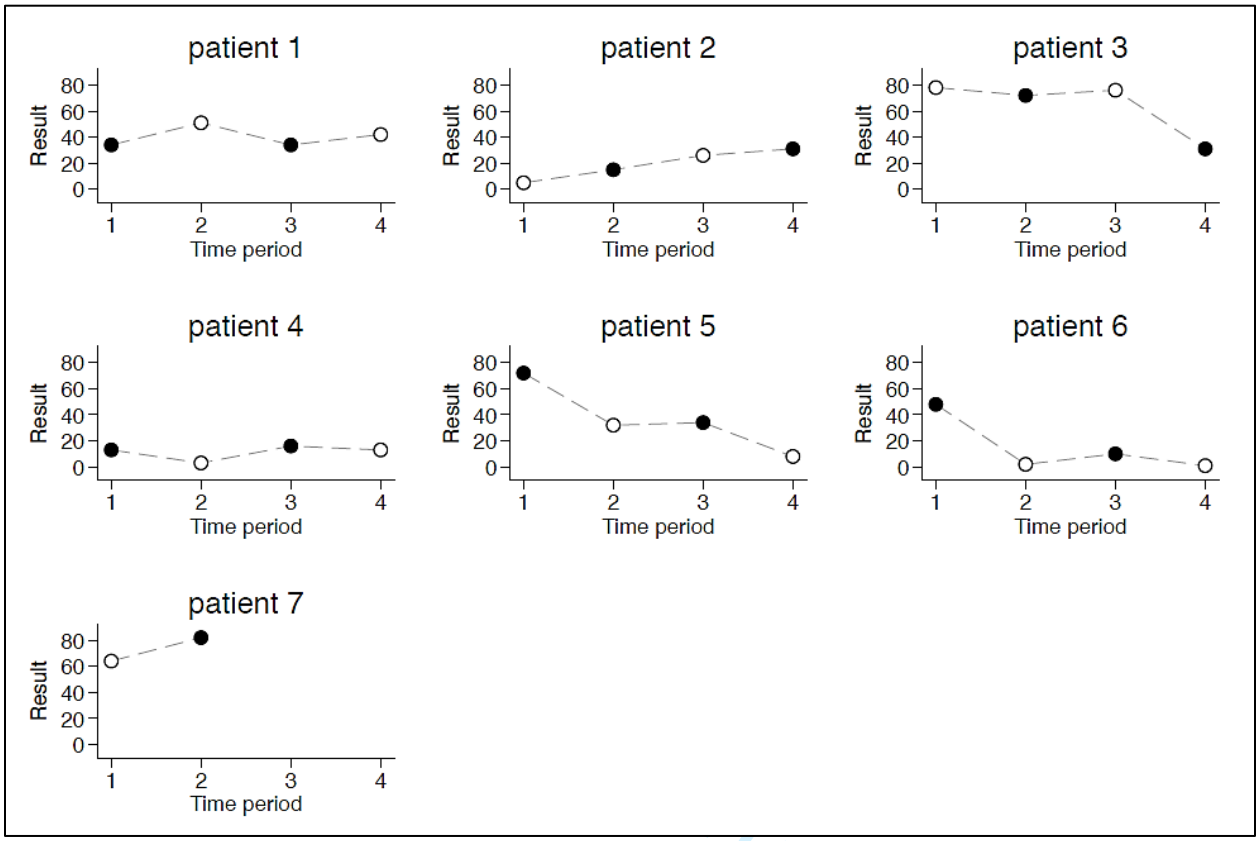
Appendix Figure 20 Legend: Data from this figure was extracted from the study published by Lashner et al in 1990, which investigates the effect of nicotine gum and placebo on abdominal pain in nonsmokers with ulcerative colitis. The average treatment effect is -3.62 (-15.84 to 8.61). White circles indicate placebo gum; black circles indicate nicotine gum.

Appendix Figure 21: Nonsmokers with ulcerative colitis treated with nicotine gum and placebo and its effect on bowel movements per day[17]



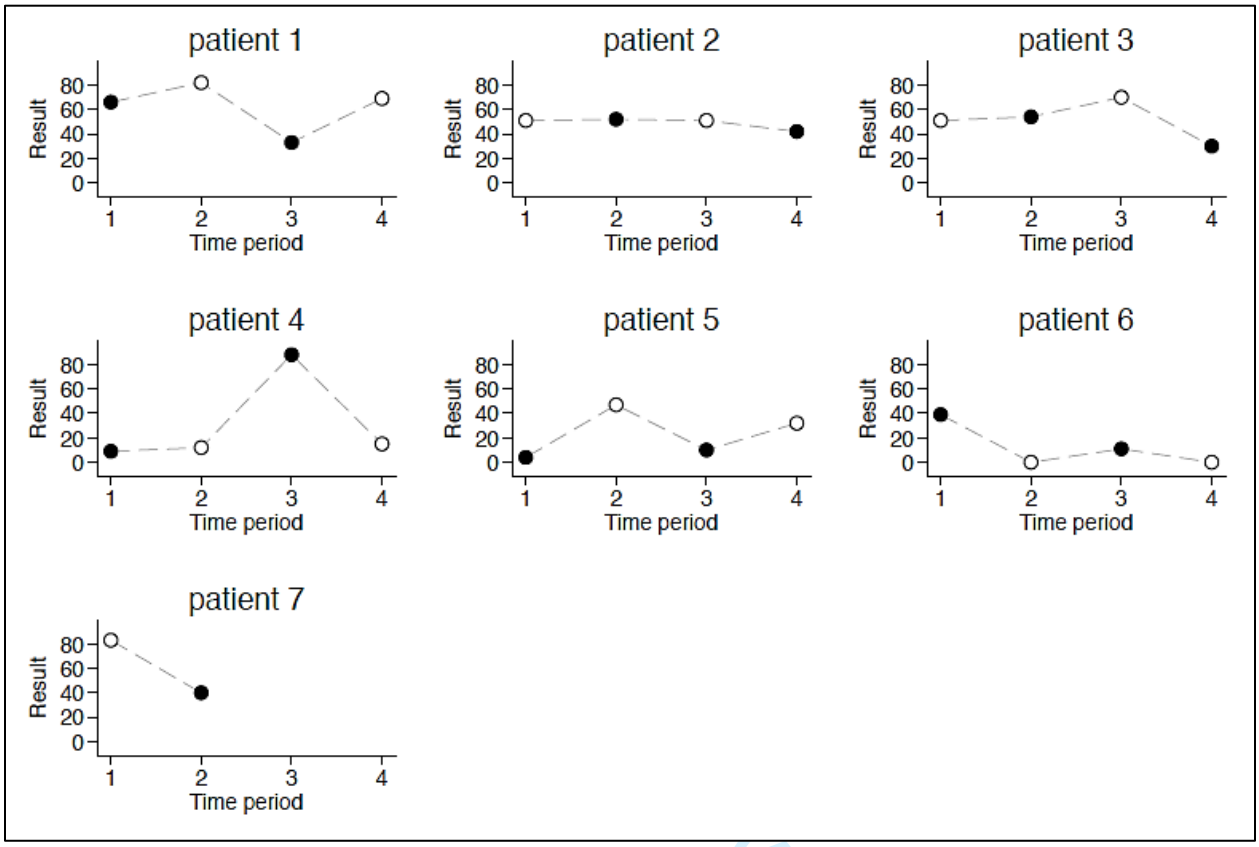
Appendix Figure 21 Legend: Data from this figure was extracted from the study published by Lashner et al in 1990, which investigates the effect of nicotine gum and placebo on bowel movements per day in nonsmokers with ulcerative colitis. The average treatment effect is -0.56 (-1.22 to 0.09). White circles indicate placebo gum; black circles indicate nicotine gum.

Appendix Figure 22: Nonsmokers with ulcerative colitis treated with nicotine gum and placebo and its effect on consistency of bowel movements[17]



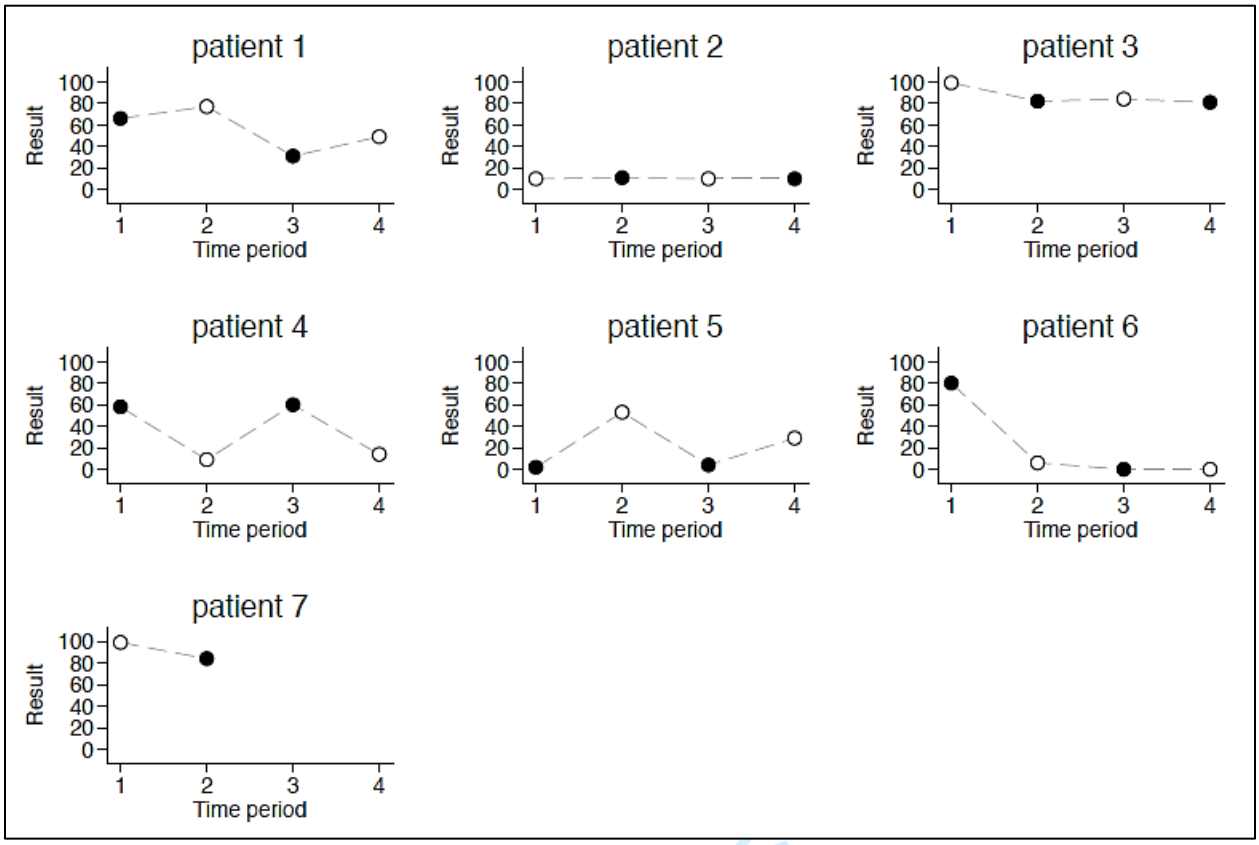
Appendix Figure 22 Legend: Data from this figure was extracted from the study published by Lashner et al in 1990, which investigates the effect of nicotine gum and placebo on consistency of bowel movements in nonsmokers with ulcerative colitis. The average treatment effect is 7.00 (-6.29 to 20.29). White circles indicate placebo gum; black circles indicate nicotine gum.

Appendix Figure 23: Nonsmokers with ulcerative colitis treated with nicotine gum and placebo and its effect on general sense of well-being[17]



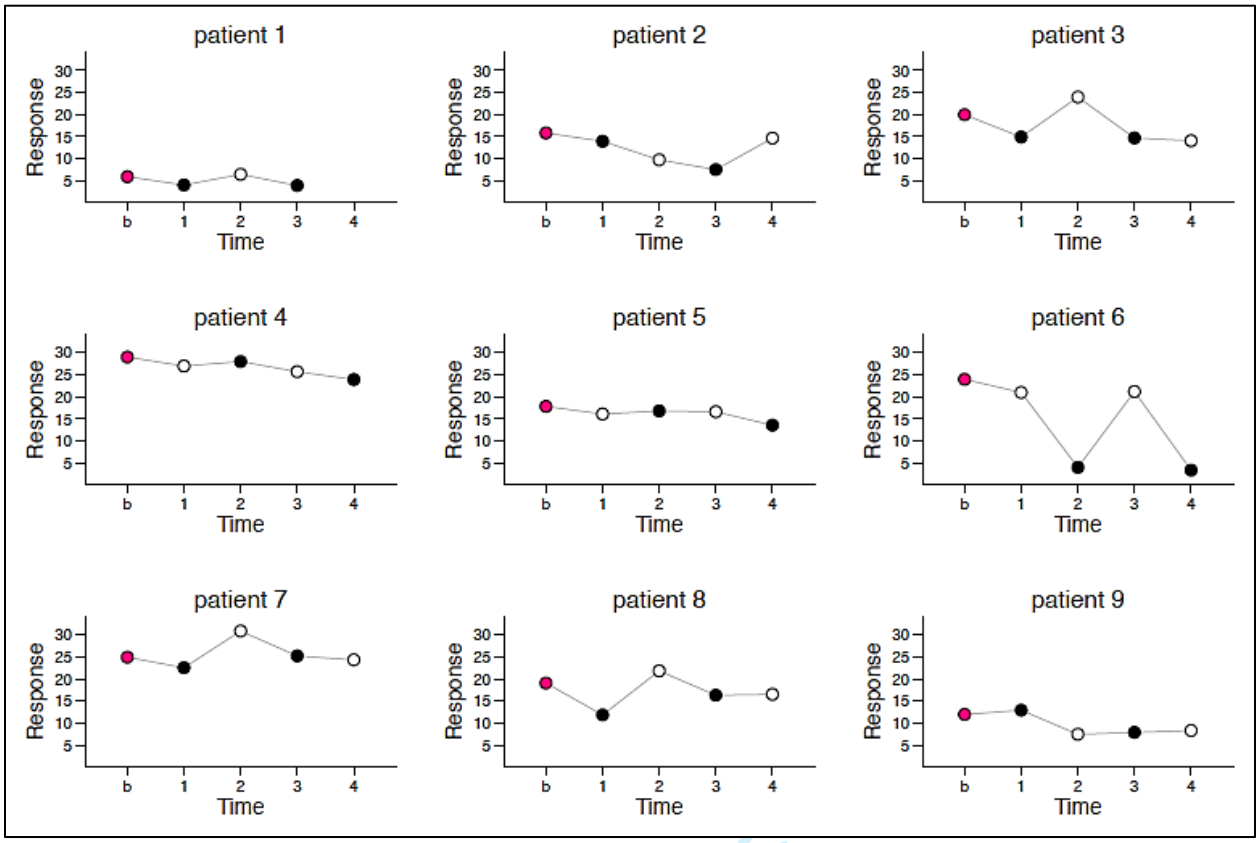
Appendix Figure 23 Legend: Data from this figure was extracted from the study published by Lashner et al in 1990, which investigates the effect of nicotine gum and placebo on general sense of well-being in nonsmokers with ulcerative colitis. The average treatment effect is -6.54 (-23.62 to 10.56). White circles indicate placebo gum; black circles indicate nicotine gum.

Appendix Figure 24: Nonsmokers with ulcerative colitis treated with nicotine gum and placebo and its effect on hematochezia[17]



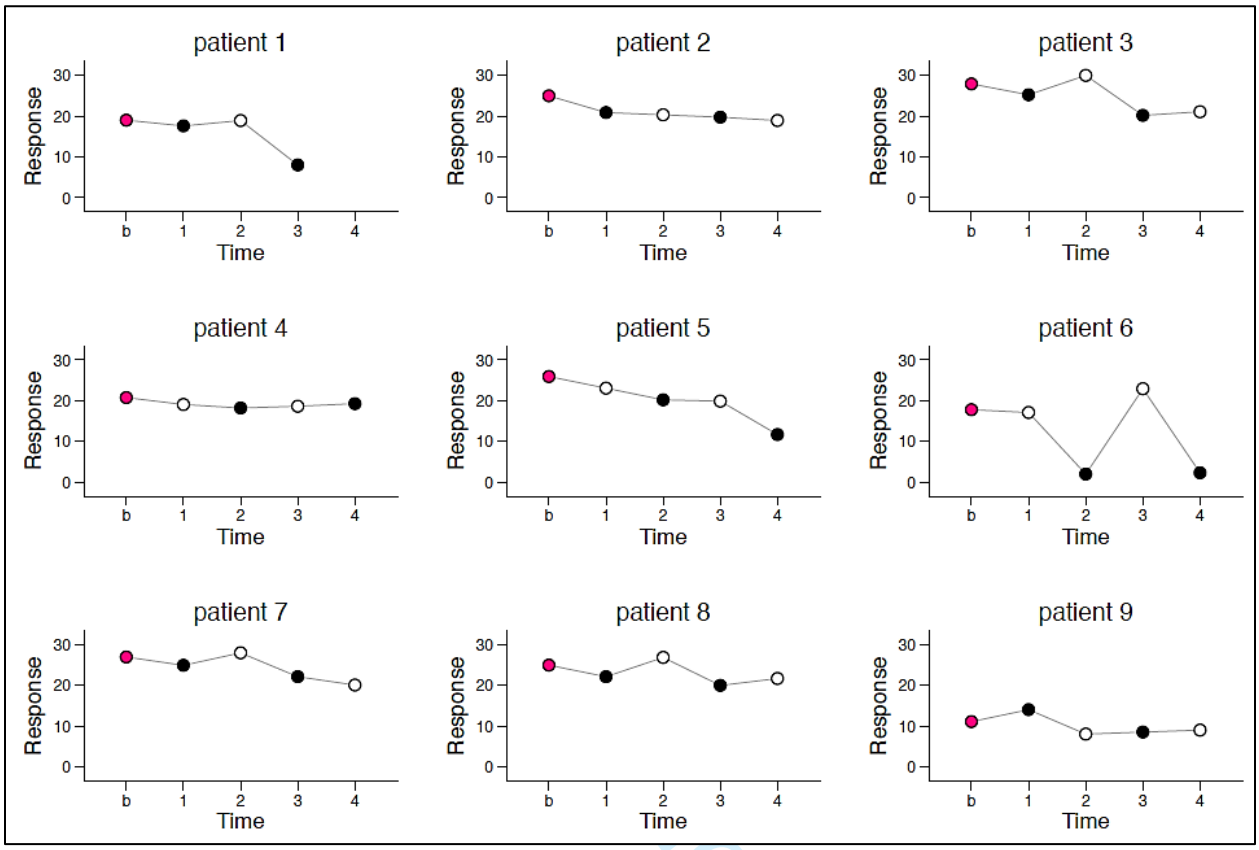
Appendix Figure 24 Legend: Data from this figure was extracted from the study published by Lashner et al in 1990, which investigates the effect of nicotine gum and placebo on hematochezia in nonsmokers with ulcerative colitis. The average treatment effect is 2.35 (-17.21 to 21.90). White circles indicate placebo gum; black circles indicate nicotine gum.

Appendix Figure 25: Patients with chronic depression and a diagnosis of major depression or dysthymia treated with sulpiride and placebo and its effect on anxiety[18]



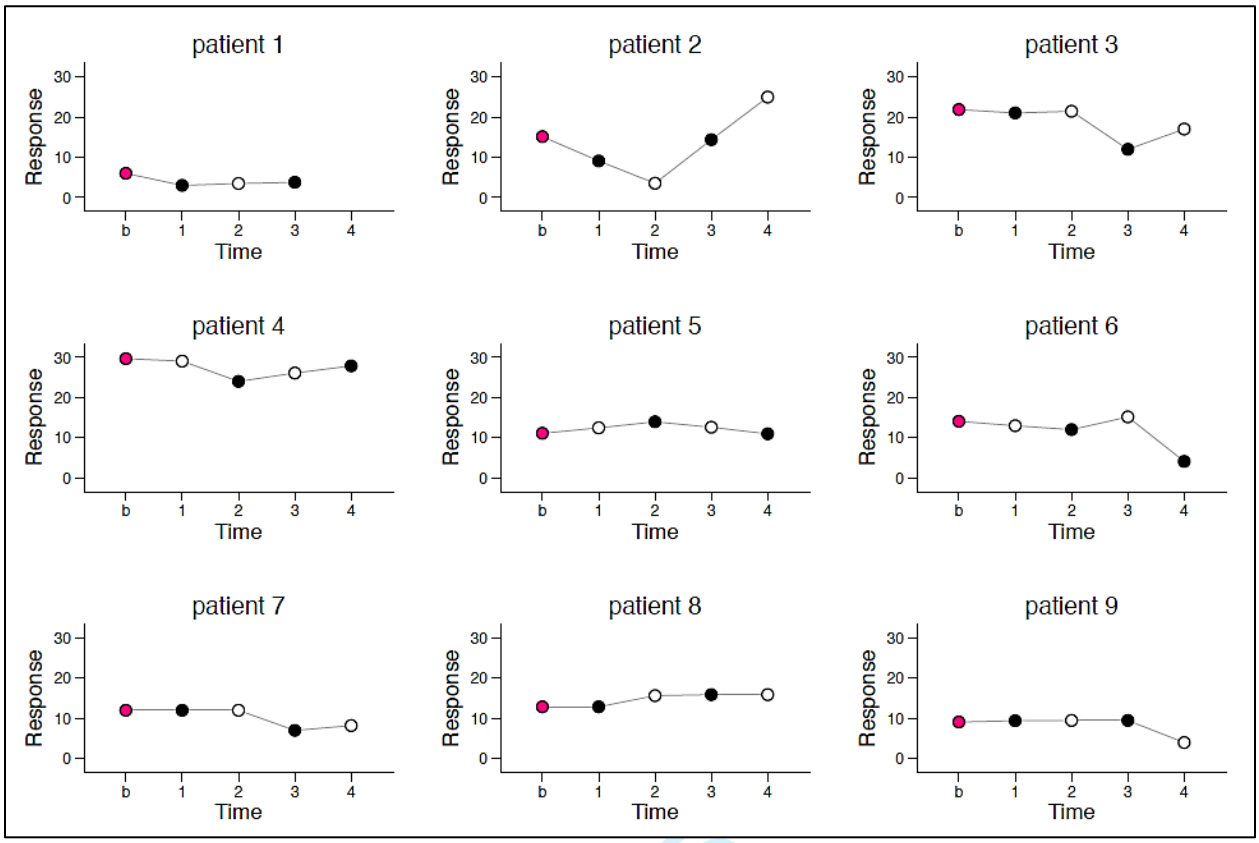
Appendix Figure 25 Legend: Data from this figure was extracted from the study published by Maier et al in 1994, which investigates the effect of sulpiride and placebo on anxiety in patients with chronic depression and a diagnosis of major depression or dysthymia. The average treatment effect is -3.81 (-7.22 to -0.40). Red circles indicate baseline; white circles indicate placebo; black circles indicate sulpiride.

Appendix Figure 26: Patients with chronic depression and a diagnosis of major depression or dysthymia treated with sulpiride and placebo and its effect on depressed mood[18]



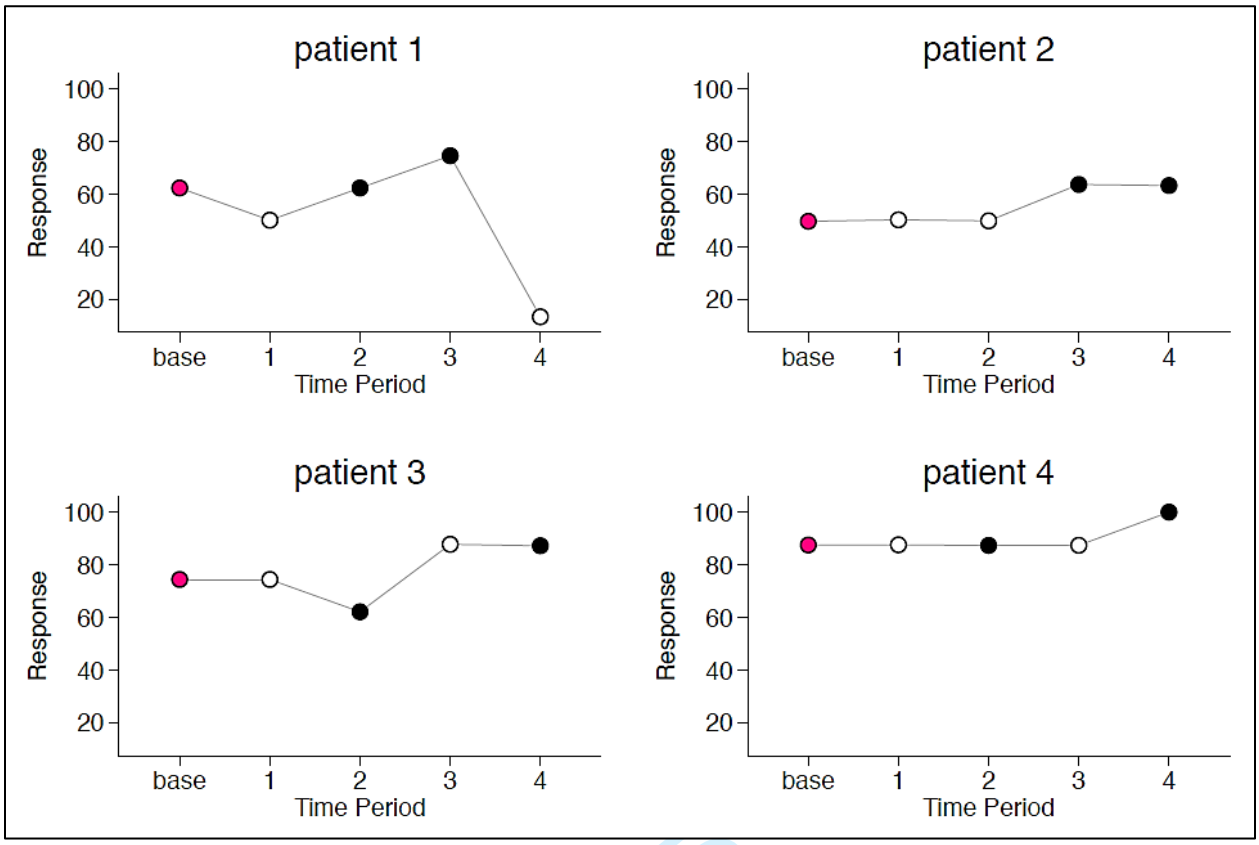
Appendix Figure 26 Legend: Data from this figure was extracted from the study published by Maier et al in 1994, which investigates the effect of sulpiride and placebo on depressed mood in patients with chronic depression and a diagnosis of major depression or dysthymia. The average treatment effect is -3.63 (-7.40 to 0.15). Red circles indicate baseline; white circles indicate placebo; black circles indicate sulpiride.

Appendix Figure 27: Patients with chronic depression and a diagnosis of major depression or dysthymia treated with sulpiride and placebo and its effect on somatization[18]



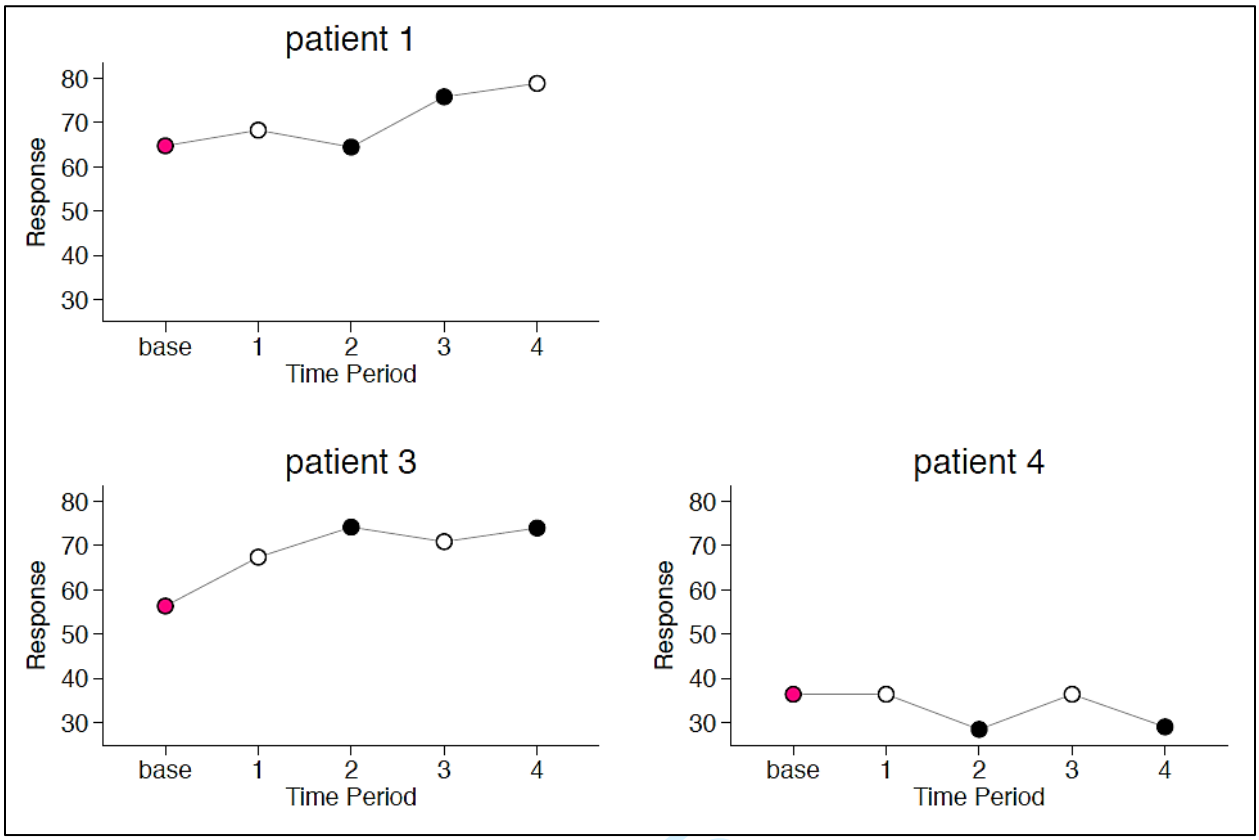
Appendix Figure 27 Legend: Data from this figure was extracted from the study published by Maier et al in 1994, which investigates the effect of sulpiride and placebo on somatization in patients with chronic depression and a diagnosis of major depression or dysthymia. The average treatment effect is -1.50 (-4.20 to 1.21). Red circles indicate baseline; white circles indicate placebo; black circles indicate sulpiride.

Appendix Figure 28: Patients with ataxia from traumatic brain injury treated with ondansetron and placebo and its effect on lower extremity ataxia[19]



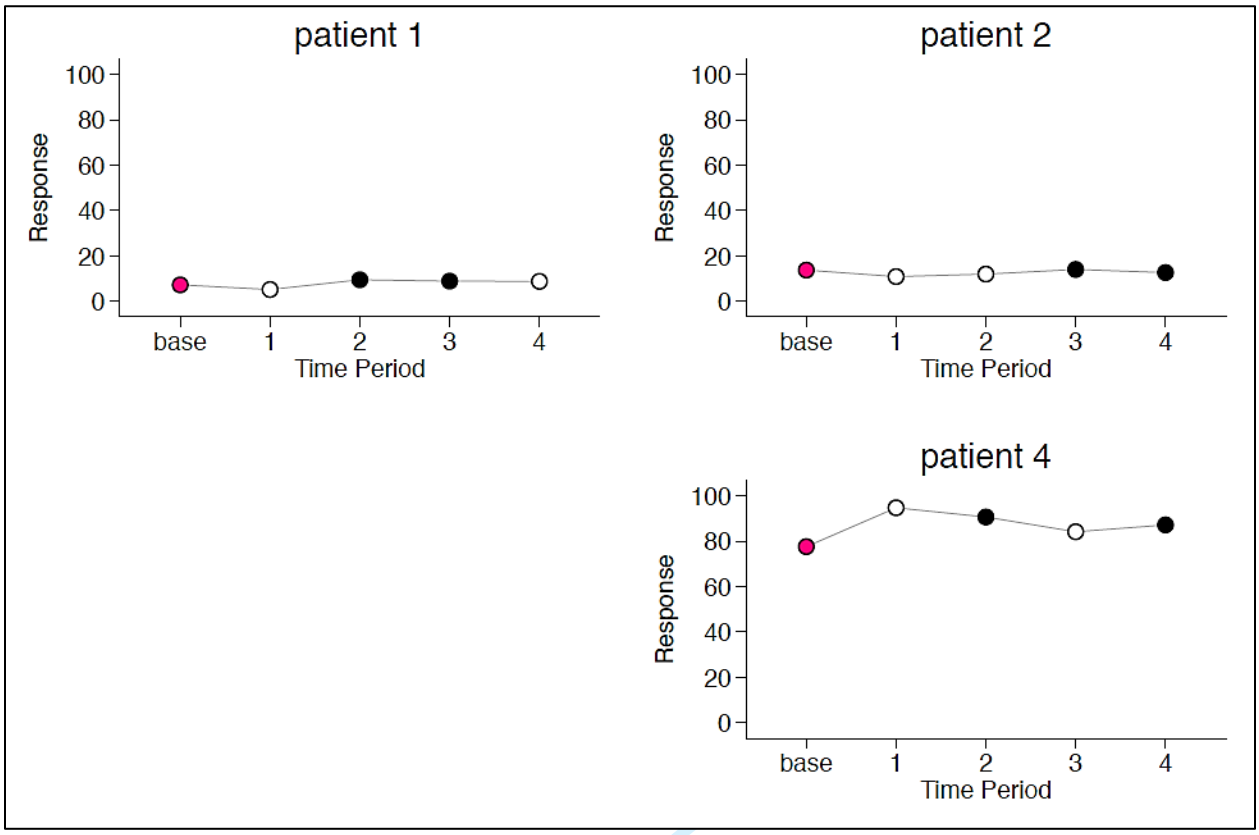
Appendix Figure 28 Legend: Data from this figure was extracted from the study published by Mandelcorn et al in 2004, which investigates the effect of ondansetron and placebo on lower extremity ataxia in patients with ataxia from traumatic brain injury. Each patient received the same treatment. The average treatment effect is 12.49 (-0.85 to 25.84). Red circles indicate baseline; white circles indicate placebo; black circles indicate ondansetron.

Appendix Figure 29: Patients with ataxia from traumatic brain injury treated with ondansetron and placebo and its effect on self-assessment score[19]



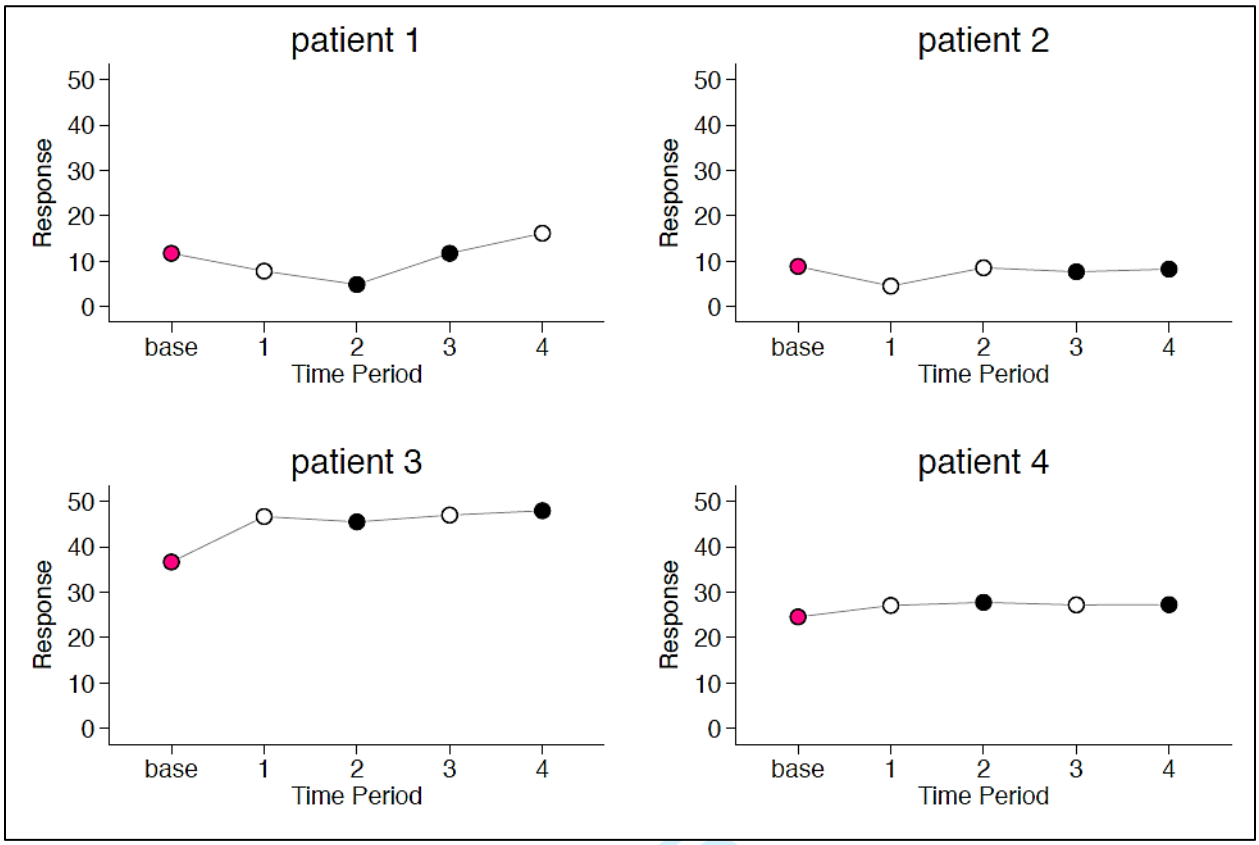
Appendix Figure 29 Legend: Data from this figure was extracted from the study published by Mandelcorn et al in 2004, which investigates the effect of ondansetron and placebo on self-assessment score in patients with ataxia from traumatic brain injury. The average treatment effect is -2.05 (-8.43 to 4.33). Red circles indicate baseline; white circles indicate placebo; black circles indicate ondansetron.

Appendix Figure 30: Patients with ataxia from traumatic brain injury treated with ondansetron and placebo and its effect on truncal ataxia[19]



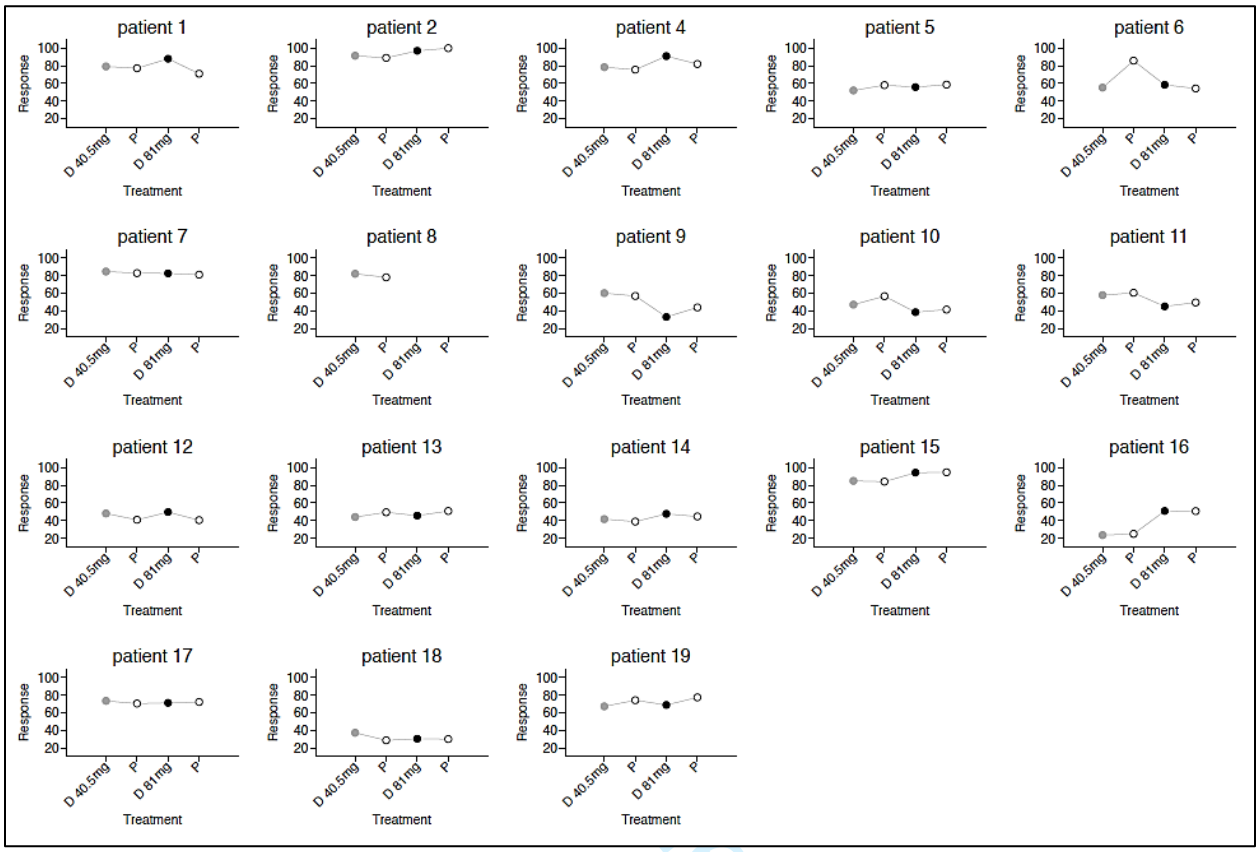
Appendix Figure 30 Legend: Data from this figure was extracted from the study published by Mandelcorn et al in 2004, which investigates the effect of ondansetron and placebo on truncal ataxia in patients with ataxia from traumatic brain injury. The average treatment effect is 1.20 (-2.06 to 4.45). Red circles indicate baseline; white circles indicate placebo; black circles indicate ondansetron.

Appendix Figure 31: Patients with ataxia from traumatic brain injury treated with ondansetron and placebo and its effect on upper extremity ataxia[19]



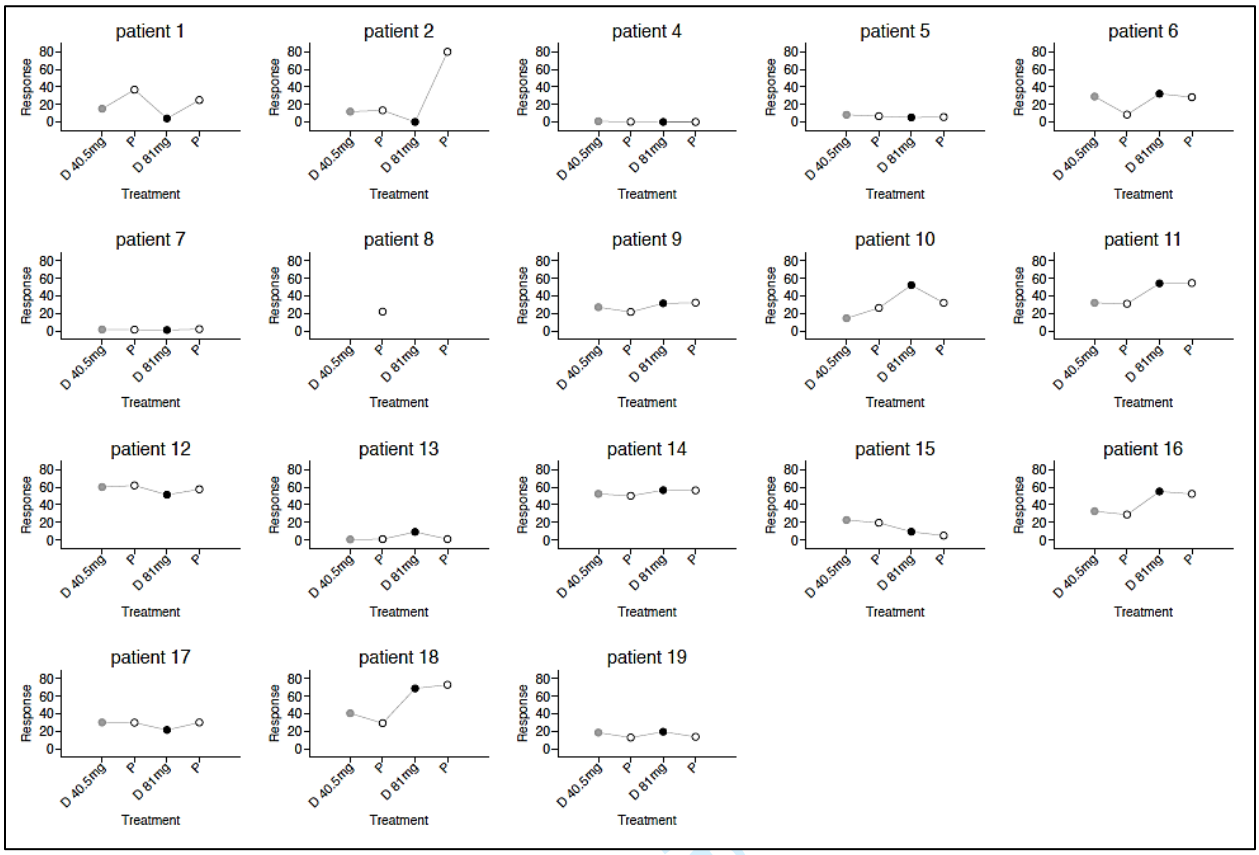
Appendix Figure 31 Legend: Data from this figure was extracted from the study published by Mandelcorn et al in 2004, which investigates the effect of ondansetron and placebo on upper extremity ataxia in patients with ataxia from traumatic brain injury. The average treatment effect is -0.50 (-3.10 to 2.10). Red circles indicate baseline; white circles indicate placebo; black circles indicate ondansetron.

Appendix Figure 32: Patients with chronic neuropathic pain treated with oral dextromethorphan and placebo and its effect on VAS pain intensity[20]



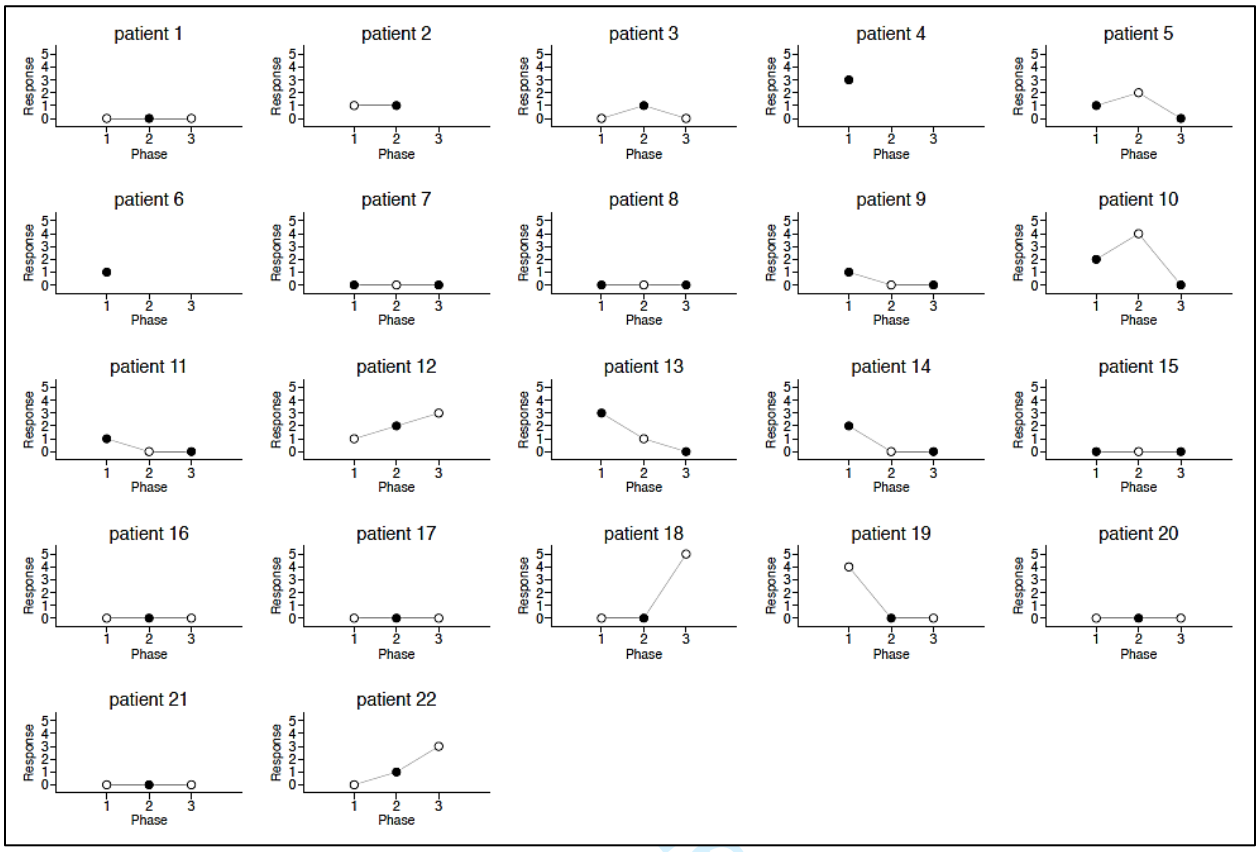
Appendix Figure 32 Legend: Data from this figure was extracted from the study published by McQuay et al in 1994, which investigates the effect of oral dextromethorphan and placebo on VAS pain intensity in patients with chronic neuropathic pain. The average treatment effect is -1.06 (-5.16 to 3.04). Grey circles indicate dextromethorphan 40.5 mg daily; black circles indicate dextromethorphan 81 mg daily; white circles indicate placebo.

Appendix Figure 33: Patients with chronic neuropathic pain treated with oral dextromethorphan and placebo and its effect on VAS relief intensity[20]



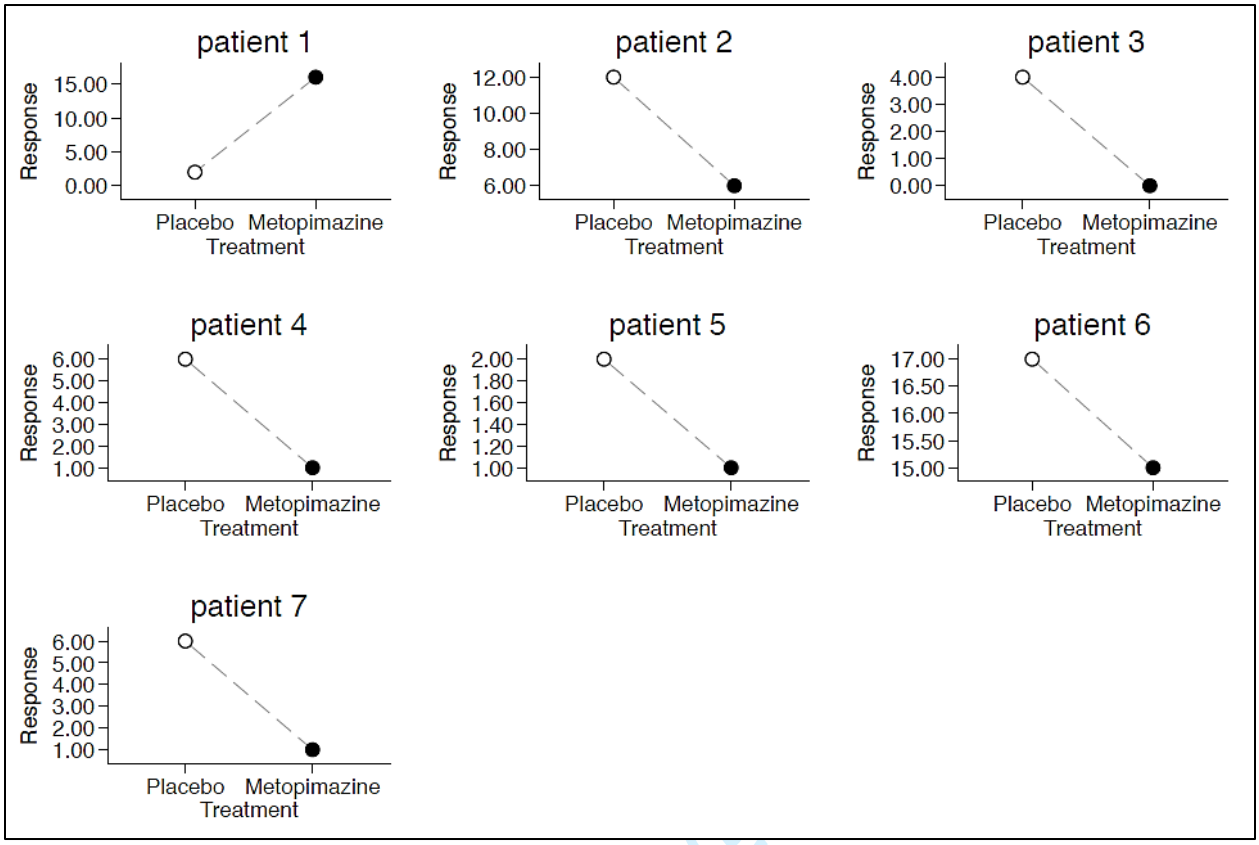
Appendix Figure 33 Legend: Data from this figure was extracted from the study published by McQuay et al in 1994, which investigates the effect of oral dextromethorphan and placebo on VAS relief intensity in patients with chronic neuropathic pain. The average treatment effect is -3.86 (-11.11 to 3.40). Grey circles indicate dextromethorphan 40.5 mg daily; black circles indicate dextromethorphan 81 mg daily; white circles indicate placebo.

Appendix Figure 34: Patients with unstable angina at rest treated with continuous and intermittent injection of isosorbide dinitrate and its effect on incidence of angina[21]



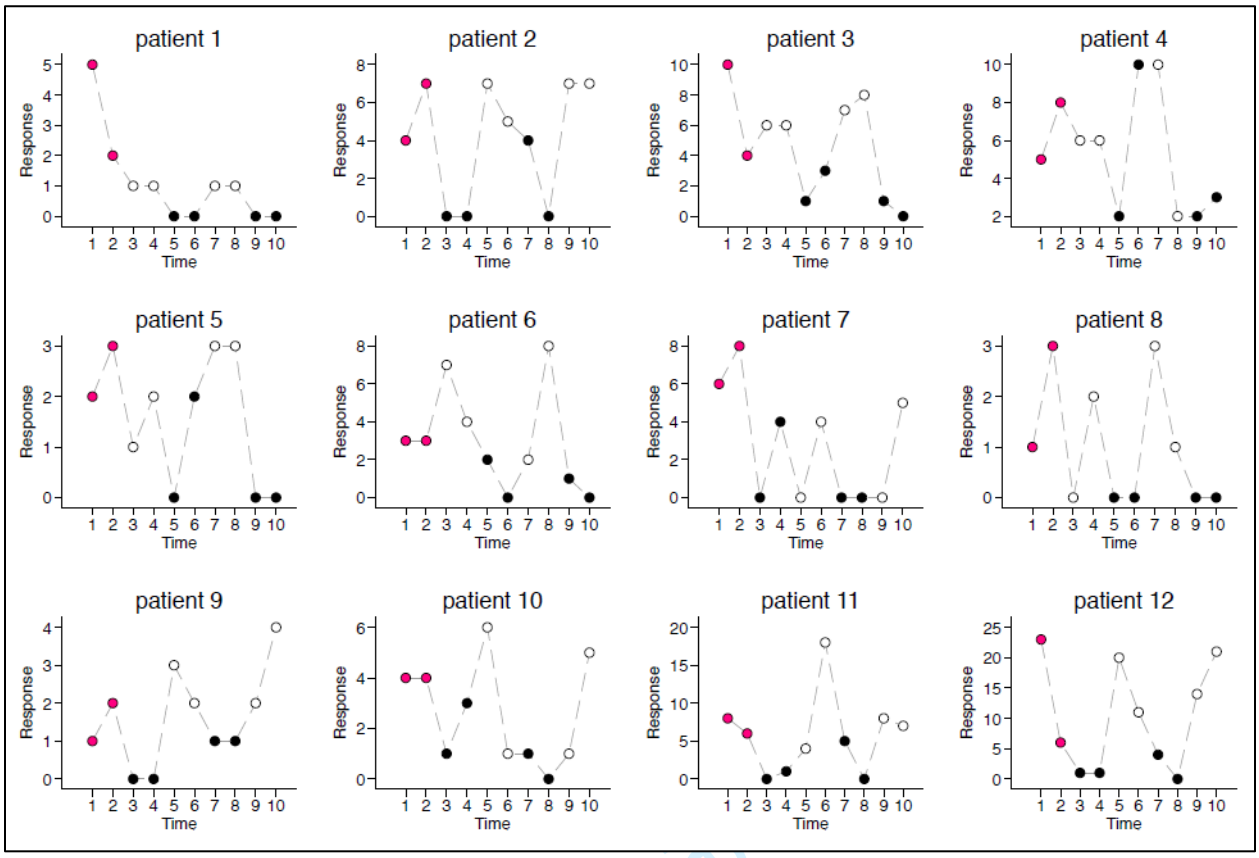
Appendix Figure 34 Legend: Data from this figure was extracted from the study published by Miyazaki et al in 1995, which investigates the effect of continuous and intermittent injection of isosorbide dinitrate on incidence of angina in patients with unstable angina. The average treatment effect is 0.47 (-0.32 to 1.26). White circles indicate continuous injection; black circles indicate intermittent injection.

Appendix Figure 35: Children with brain tumors receiving highly emetogenic therapy treated with ondansetron/metopimazine and ondansetron monotherapy and its effect on emetic episodes per day[22]



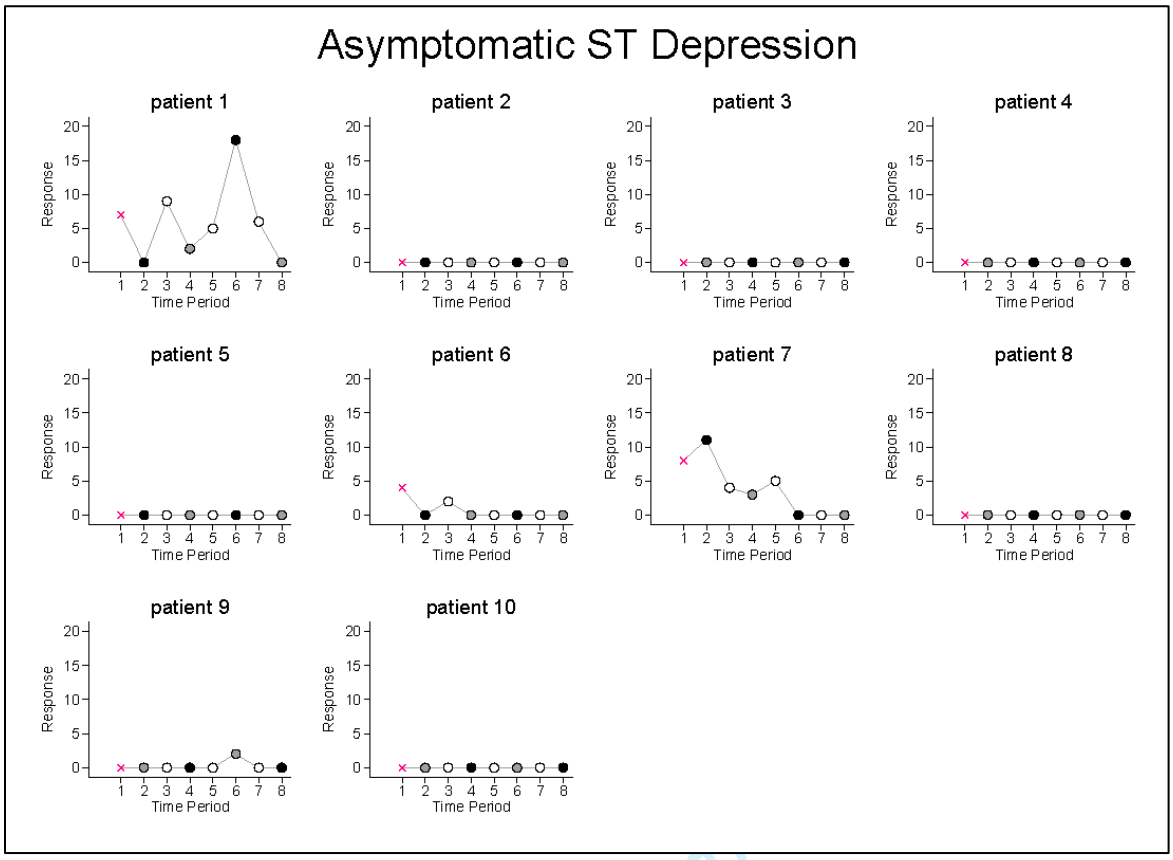
Appendix Figure 35 Legend: Data from this figure was extracted from the study published by Nathan et al in 2006, which investigates the effect of ondansetron/metopimazine and ondansetron monotherapy on emetic episodes per day in children with brain tumors receiving highly emetogenic therapy. The average treatment effect is -0.56 (-1.74 to 0.62). White circles indicate placebo; black circles indicate metopimazine.

Appendix Figure 36: Patients with unstable angina at rest treated with oral verapamil and placebo and its effect on ischemic attacks[23]



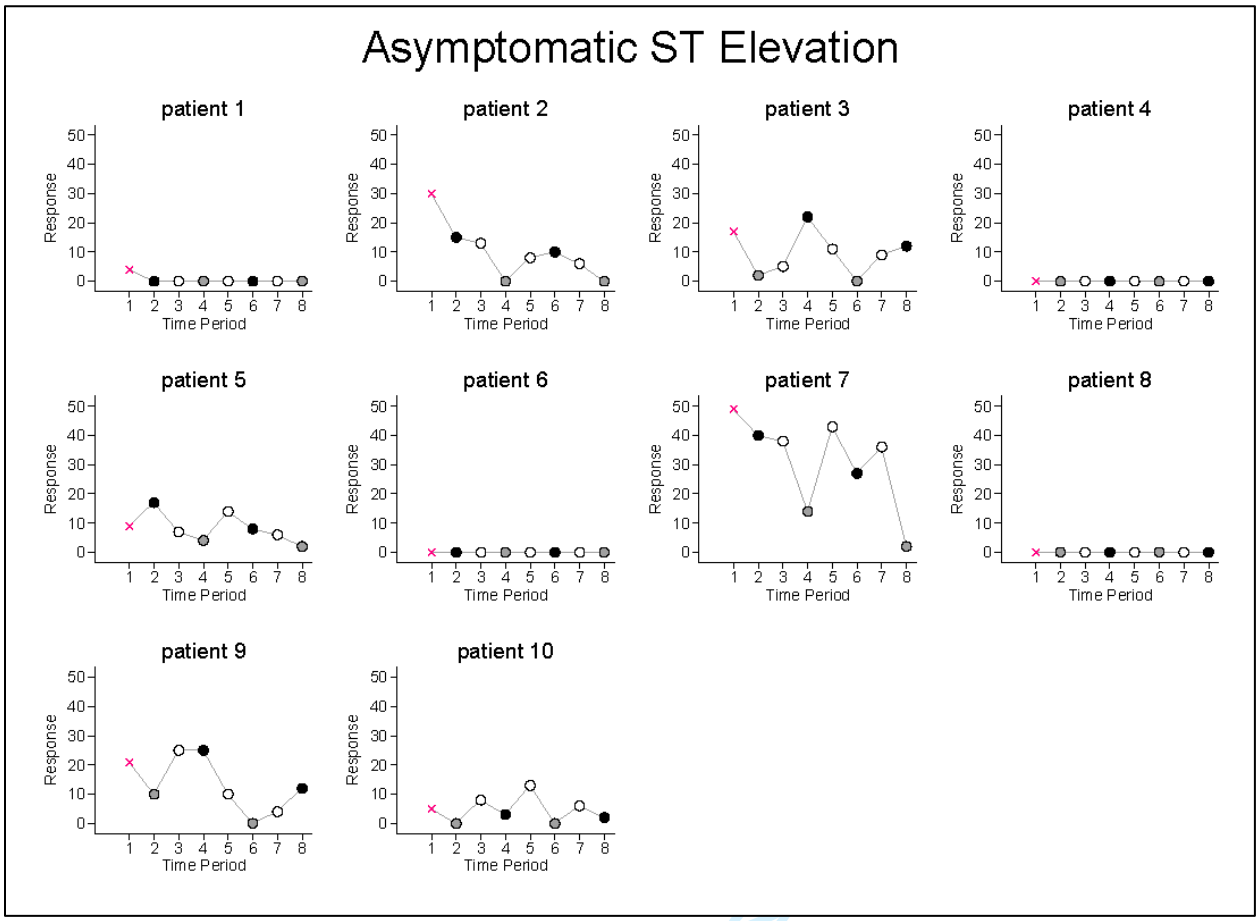
Appendix Figure 36 Legend: Data from this figure was extracted from the study published by Parodi et al in 1979, which investigates the effect of oral verapamil and placebo on ischemic attacks in patients with unstable angina. The average treatment effect is -1.63 (-2.10 to -1.17). Red circles indicate baseline; white circles indicate placebo; black circles indicate verapamil.

Appendix Figure 37: Patients with unstable angina at rest treated with verapamil, propranolol and placebo and its effect on asymptomatic ST depression[24]



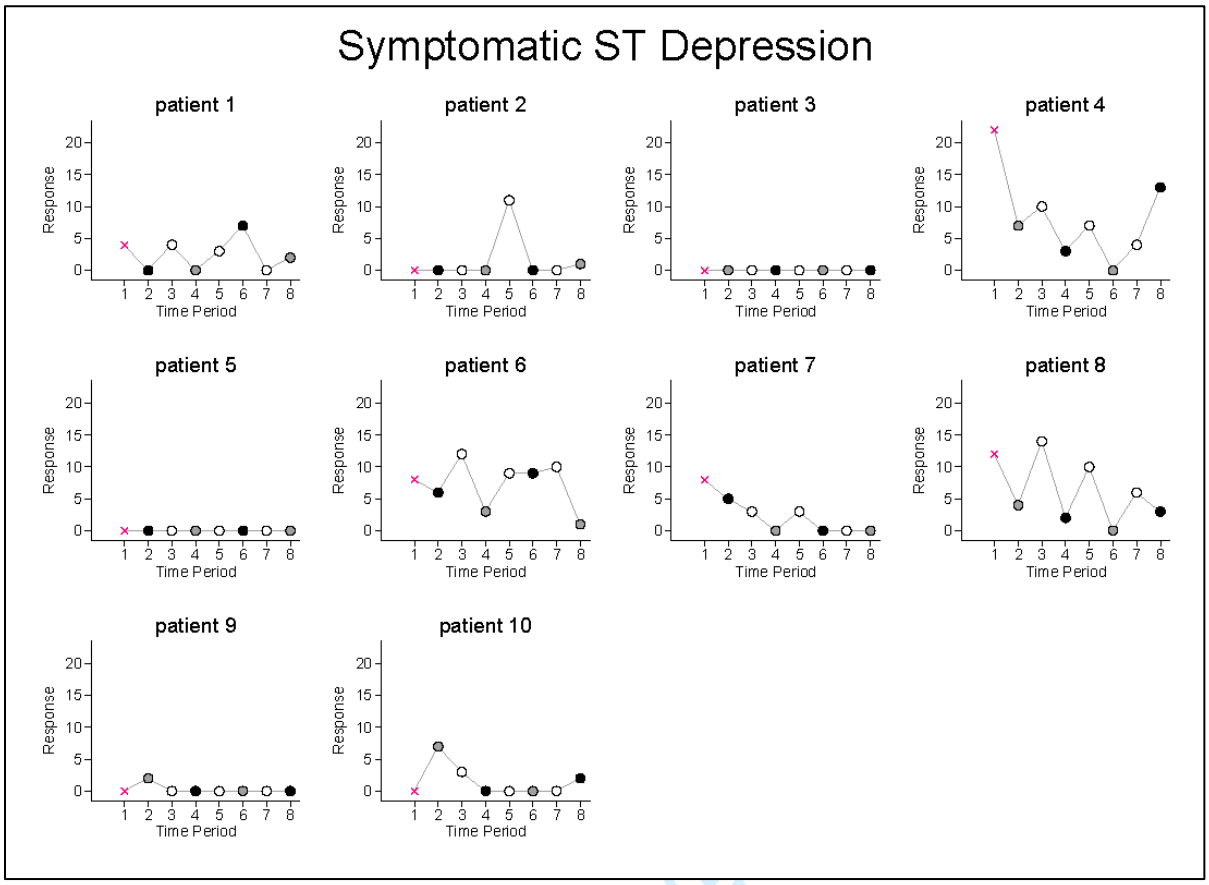
Appendix Figure 37 Legend: Data from this figure was extracted from the study published by Parodi et al in 1986, which investigates the effect of verapamil, propranolol and placebo on asymptomatic ST depression in patients with unstable angina. The average treatment effect is -0.82 (-2.54 to 0.90). Red Xs indicate baseline; white circles indicate placebo; grey circles indicate propranolol; black circles indicate verapamil.

Appendix Figure 38: Patients with unstable angina at rest treated with verapamil, propranolol and placebo and its effect on asymptomatic ST elevation[24]



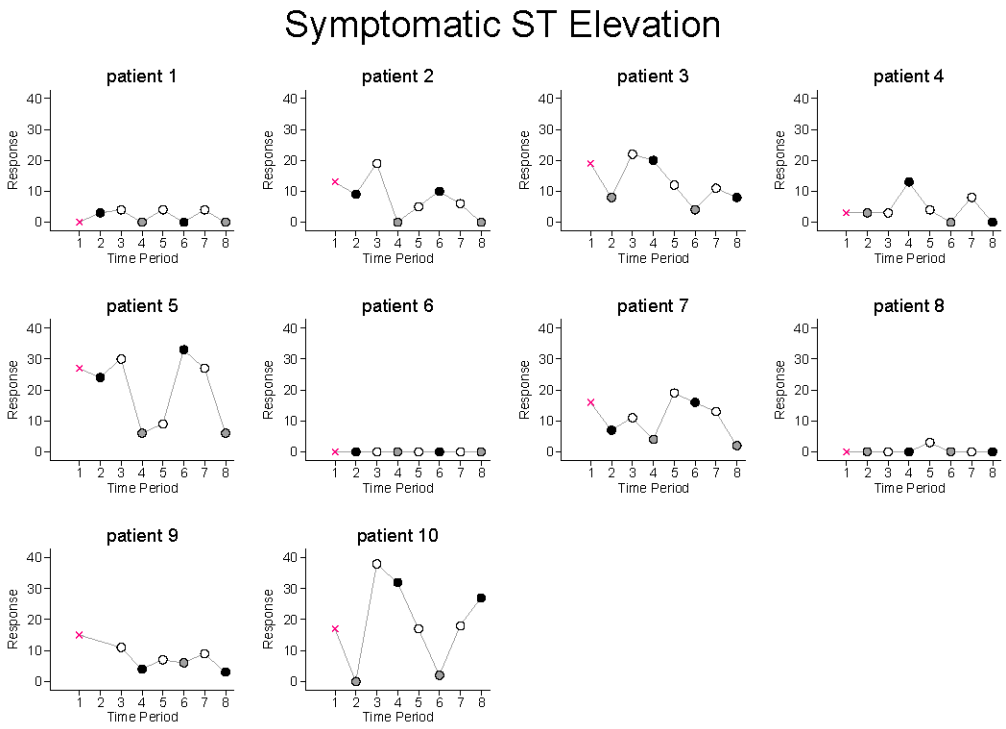
Appendix Figure 38 Legend: Data from this figure was extracted from the study published by Parodi et al in 1986, which investigates the effect of verapamil, propranolol and placebo on asymptomatic ST elevation in patients with unstable angina. The average treatment effect is -1.97 (-2.92 to -1.01). Red Xs indicate baseline; white circles indicate placebo; grey circles indicate propranolol; black circles indicate verapamil.

Appendix Figure 39: Patients with unstable angina at rest treated with verapamil, propranolol and placebo and its effect on symptomatic ST depression[24]



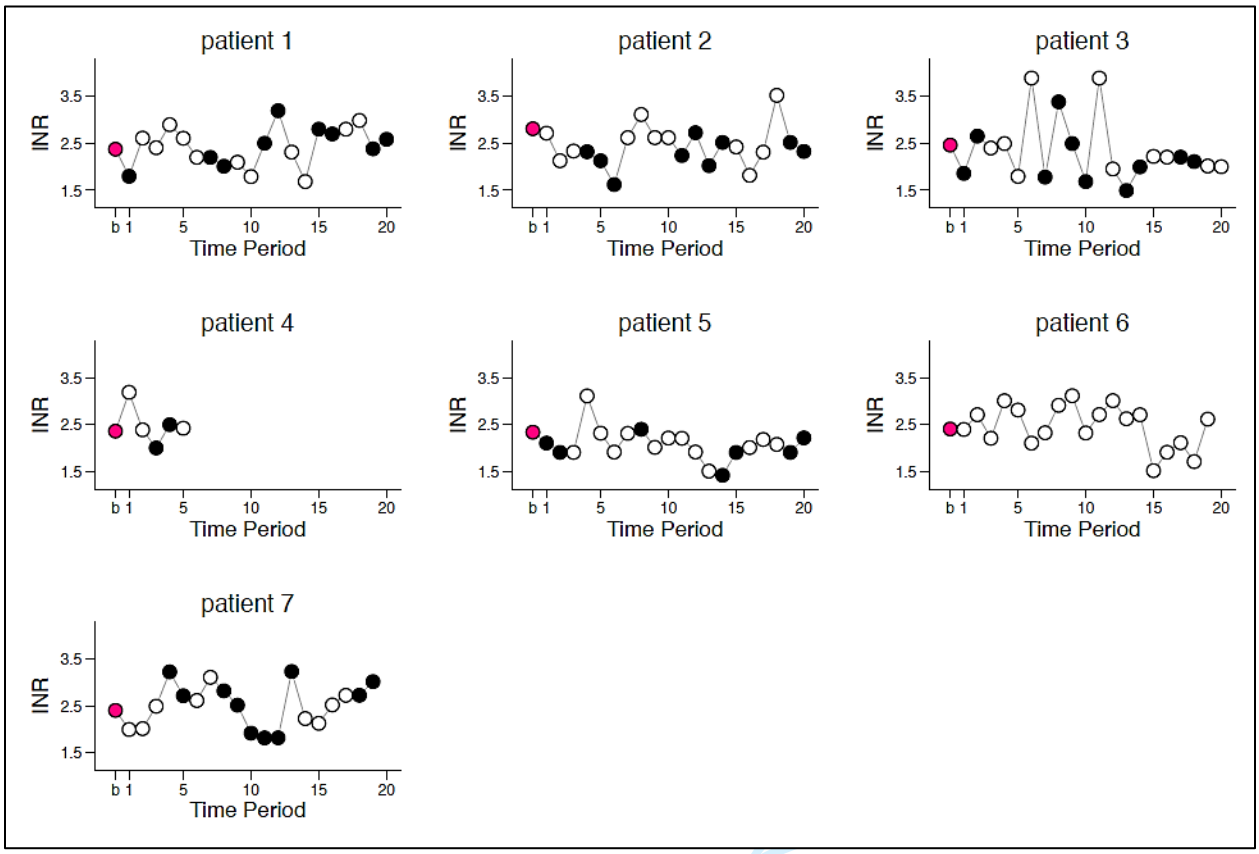
Appendix Figure 39 Legend: Data from this figure was extracted from the study published by Parodi et al in 1986, which investigates the effect of verapamil, propranolol and placebo on symptomatic ST depression in patients with unstable angina. The average treatment effect is -0.98 (-1.84 to -0.13). Red Xs indicate baseline; white circles indicate placebo; grey circles indicate propranolol; black circles indicate verapamil.

Appendix Figure 40: Patients with unstable angina at rest treated with verapamil, propranolol and placebo and its effect on symptomatic ST elevation[24]



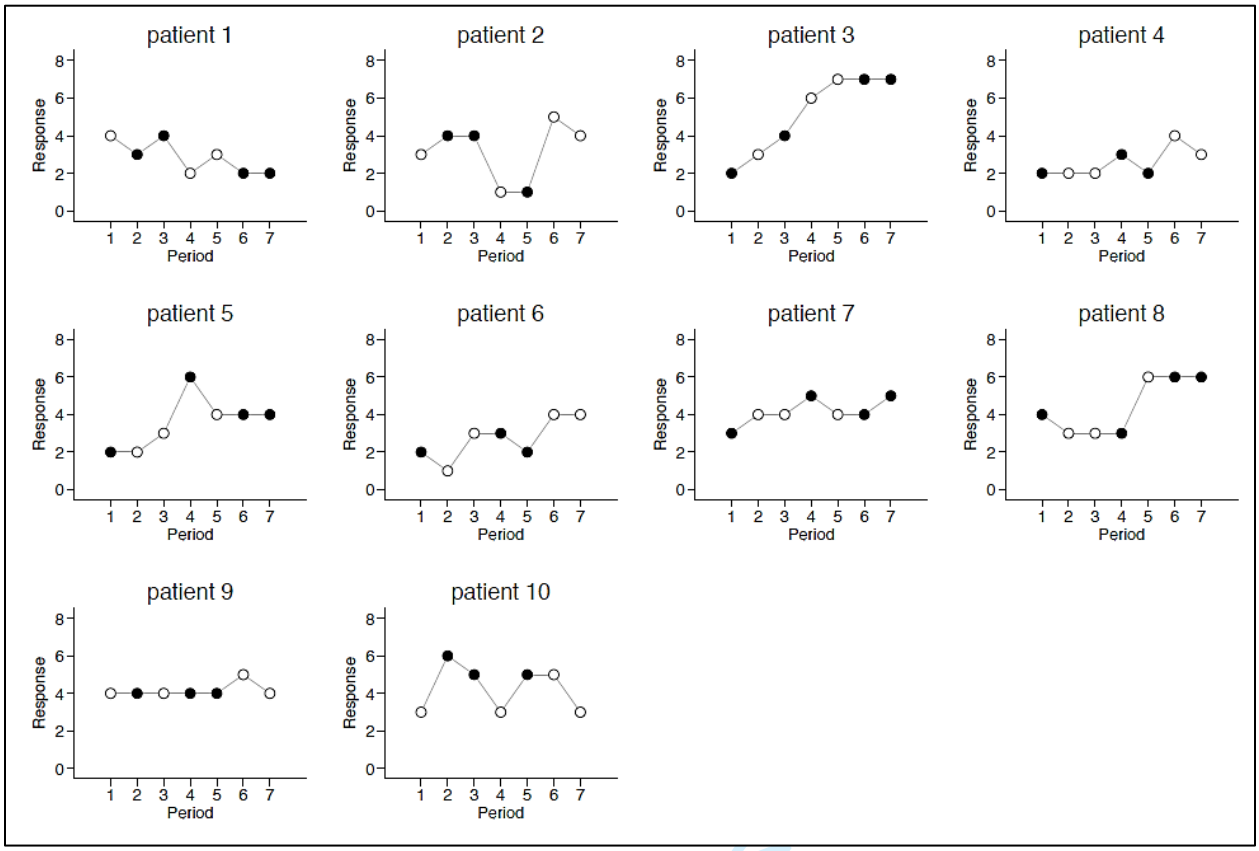
Appendix Figure 40 Legend: Data from this figure was extracted from the study published by Parodi et al in 1986, which investigates the effect of verapamil, propranolol and placebo on symptomatic ST elevation in patients with unstable angina. The average treatment effect is -1.87 (-2.72 to -1.02). Red Xs indicate baseline; white circles indicate placebo; grey circles indicate propranolol; black circles indicate verapamil.

Appendix Figure 41: Patients previously taking warfarin for either atrial fibrillation or deep vein thrombosis treated with apo-warfarin and coumadin and its effect on international normalized ratio[8]



Appendix Figure 41 Legend: Data from this figure was extracted from the study published by Pereira et al in 1995, which investigates the effect of apo-warfarin and coumadin on international normalized ratio in patients previously taking warfarin for either atrial fibrillation or deep vein thrombosis. The average treatment effect is -0.12 (-0.30 to 0.07). Red circles indicate baseline; white circles indicate Coumadin; black circles indicate apo-warfarin.

Appendix Figure 42: Parkinson's disease patients with troublesome dyskinesia treated with simvastatin and placebo and its effect on discomfort caused by troublesome dyskinesia[25]



Appendix Figure 42 Legend: Data from this figure was extracted from the study published by Tison et al in 2012, which investigates the effect of simvastatin and placebo on discomfort caused by troublesome dyskinesia in Parkinson's disease patients with troublesome dyskinesia. The average treatment effect is0.20 (-0.40 to 0.80). White circles indicate placebo; black circles indicate simvastatin.

Appendix Reference List

(1) Emmanuel AV, Kamm MA, Roy AJ, Kerstens R, Vandeplassche L. Randomised clinical trial: the efficacy of prucalopride in patients with chronic intestinal pseudo-obstruction--a double-blind, placebo-controlled, cross-over, multiple n = 1 study. *Aliment Pharmacol Ther* 2012; 35(1):48-55.

(2) Haas DC, Sheehe PR. Dextroamphetamine pilot crossover trials and n of 1 trials in patients with chronic tension-type and migraine headache. *Headache* 2004; 44(10):1029-1037.

(3) Jaeschke R, Adachi J, Guyatt G, Keller J, Wong B. Clinical usefulness of amitriptyline in fibromyalgia: the results of 23 N-of-1 randomized controlled trials. *J Rheumatol* 1991; 18(3):447-451.

(4) Johannessen T, Petersen H, Kristensen P, Fosstvedt D, Kleveland PM, Dybdahl J et al. Cimetidine on-demand in dyspepsia. Experience with randomized controlled single-subject trials. *Scand J Gastroenterol* 1992; 27(3):189-195.

(5) Mahon J, Laupacis A, Donner A, Wood T. Randomised study of n of 1 trials versus standard practice. *BMJ* 1996; 312(7038):1069-1074.

(6) March L, Irwig L, Schwarz J, Simpson J, Chock C, Brooks P. n of 1 trials comparing a non-steroidal anti-inflammatory drug with paracetamol in osteoarthritis. *BMJ* 1994; 309(6961):1041-1045.

(7) Patel A, Jaeschke R, Guyatt GH, Keller JL, Newhouse MT. Clinical usefulness of n-of-1 randomized controlled trials in patients with nonreversible chronic airflow limitation. *Am Rev Respir Dis* 1991; 144(4):962-964.

(8) Pereira JA, Holbrook AM, Dolovich L, Goldsmith C, Thabane L, Douketis JD et al. Are brand-name and generic warfarin interchangeable? Multiple n-of-1 randomized, crossover trials. *Ann Pharmacother* 2005; 39(7-8):1188-1193.

(9) Wallace AE, Kofoed LL. Statistical analysis of single case studies in the clinical setting: the example of methylphenidate trials in children with attention-deficit hyperactivity disorder. *J Child Adolesc Psychopharmacol* 1994; 4(3):141-150.

(10) Woodfield R, Goodyear-Smith F, Arroll B. N-of-1 trials of quinine efficacy in skeletal muscle cramps of the leg. *Br J Gen Pract* 2005; 55(512):181-185.

(11) Zucker DR, Ruthazer R, Schmid CH, Feuer JM, Fischer PA, Kieval RI et al. Lessons learned combining N-of-1 trials to assess fibromyalgia therapies. *J Rheumatol* 2006; 33(10):2069-2077.

(12) Joy TR, Monjed A, Zou GY, Hegele RA, McDonald CG, Mahon JL. N-of-1 (single-patient) trials for statin-related myalgia. *Ann Intern Med* 2014; 160(5):301-310.

(13) Lipka AF, Vrinten C, van Zwet EW, Schimmel KJ, Cornel MC, Kuijpers MR et al. Ephedrine treatment for autoimmune myasthenia gravis. *Neuromuscul Disord* 2017; 27(3):259-265.

(14) Camfield P, Gordon K, Dooley J, Camfield C. Melatonin appears ineffective in children with intellectual deficits and fragmented sleep: six "N of 1" trials. *J Child Neurol* 1996; 11(4):341-343.

(15) Hinderer SR. The supraspinal anxiolytic effect of baclofen for spasticity reduction. *Am J Phys Med Rehabil* 1990; 69(5):254-258.

(16) Langer JC, Winthrop AL, Issenman RM. The single-subject randomized trial. A useful clinical tool for assessing therapeutic efficacy in pediatric practice. *Clin Pediatr (Phila)* 1993; 32(11):654-657.

(17) Lashner BA, Hanauer SB, Silverstein MD. Testing nicotine gum for ulcerative colitis patients. Experience with single-patient trials. *Dig Dis Sci* 1990; 35(7):827-832.

(18) Maier W, Benkert O. Treatment of chronic depression with sulpiride: evidence of efficacy in placebo-controlled single case studies. *Psychopharmacology (Berl)* 1994; 115(4):495-501.

(19) Mandelcorn J, Cullen NK, Bayley MT. A preliminary study of the efficacy of ondansetron in the treatment of ataxia, poor balance and incoordination from brain injury. *Brain Inj* 2004; 18(10):1025-1039.

(20) McQuay HJ, Carroll D, Jadad AR, Glynn CJ, Jack T, Moore RA et al. Dextromethorphan for the treatment of neuropathic pain: a double-blind randomised controlled crossover trial with integral n-of-1 design. *Pain* 1994; 59(1):127-133.

(21) Miyazaki S, Nonogi H, Goto Y, Sumiyoshi T, Haze K, Hiramori K. Comparison of the therapeutic efficacy of continuous and intermittent injection of isosorbide dinitrate: a randomized study on unstable angina. *Intern Med* 1995; 34(9):856-862.

(22) Nathan PC, Tomlinson G, Dupuis LL, Greenberg ML, Ota S, Bartels U et al. A pilot study of ondansetron plus metopimazine vs. ondansetron monotherapy in children receiving highly emetogenic chemotherapy: a Bayesian randomized serial N-of-1 trials design. *Support Care Cancer* 2006; 14(3):268-276.

(23) Parodi O, Maseri A, Simonetti I. Management of unstable angina at rest by verapamil. A double-blind cross-over study in coronary care unit. *Br Heart J* 1979; 41(2):167-174.

(24) Parodi O, Simonetti I, Michelassi C, Carpeggiani C, Biagini A, L'Abbate A et al. Comparison of verapamil and propranolol therapy for angina pectoris at rest: a randomized, multiple-crossover, controlled trial in the coronary care unit. *Am J Cardiol* 1986; 57(11):899-906.

(25) Tison F, Negre-Pages L, Meissner WG, Dupouy S, Li Q, Thiolat ML et al. Simvastatin decreases levodopa-induced dyskinesia in monkeys, but not in a randomized, placebo-controlled, multiple cross-over ("n-of-1") exploratory trial of simvastatin against levodopa-induced dyskinesia in Parkinson's disease patients. *Parkinsonism Relat Disord* 2013; 19(4):416-421.

(26) Joy TR, Monjed A, Zou GY, Hegele RA, McDonald CG, Mahon JL. N-of-1 (single-patient) trials for statin-related myalgia. *Ann Intern Med* 2014; 160(5):301-310.

Appendix Table 5. Studies reporting person-level treatment effect with both fixed-effect and random-effect using a method of moments estimator

Study	Outcome	Fixed effect model	P for HTE (fixed-effects model)	Random Treatment Effect	summary_tau2	P for HTE (random-effects model)
March 1994[6]	Mean pain score on VAS taken from 2nd week of tx	-4.155 (-4.807 to -3.502)	<0.001	-7.093 (-11.939 to -2.248)	73.530	<0.001
March 1994[6]	Mean stiffness score on VAS taken from 2nd week of	-2.192 (-2.549 to -1.835)	<0.001	-5.992 (-11.280 to -0.704)	88.872	<0.001
Emmanuel 2012[1]	Bloating	-0.131 (-0.171 to -0.090)	<0.001	-0.344 (-0.619 to -0.069)	0.071	<0.001
Emmanuel 2012[1]	Pain	-0.160 (-0.209 to -0.111)	<0.001	-0.440 (-0.771 to -0.110)	0.106	<0.001
Haas 2004[2]	Chronic tension-type headache grade	0.733 (0.609 to 0.857)	<0.001	0.772 (0.454 to 1.090)	0.350	<0.001
Haas 2004[2]	Chronic tension-type headache grade	0.543 (0.394 to 0.693)	0.067	0.542 (0.354 to 0.731)	0.055	0.067
Jaeschke 1991[3]	7-point symptom scale	0.356 (0.286 to 0.426)	<0.001	0.427 (0.210 to 0.645)	0.186	<0.001
Jaeschke 1991[3]	Tender point changes count	1.072 (0.701 to 1.443)	<0.001	1.320 (0.404 to 2.236)	2.166	<0.001
Johannessen 1992[4]	6-point symptom scale	0.657 (0.530 to 0.785)	<0.001	0.698 (0.466 to 0.931)	0.382	<0.001
Joy 2014[26]	VAS myalgia score	0.119 (-2.283 to 2.521)	0.995	0.119 (-2.283 to 2.521)	0.000	0.996
Joy 2014[26]	Symptom-specific VAS	1.937 (0.179 to 3.696)	0.797	1.937 (0.179 to 3.696)	0.000	0.797
Joy 2014[26]	Pain severity score	0.086 (-0.215 to 0.387)	0.986	0.086 (-0.215 to 0.387)	0.000	0.986

Joy 2014[26]	Pain interference score	-0.016 (-0.095 to 0.064)	0.917	-0.016 (-0.095 to 0.064)	0.000	0.917
Lipka 2017[13]	Quantitative myasthenia gravis score	1.006 (0.215 to 1.797)	0.803	1.006 (0.215 to 1.797)	0.000	0.803
Lipka 2017[13]	Myasthenia gravis composite	2.952 (0.969 to 4.934)	0.177	2.891 (0.348 to 5.433)	2.631	0.177
Lipka 2017[13]	MG-ADL	1.110 (0.269 to 1.951)	0.047	1.099 (-0.277 to 2.474)	1.222	0.047
Lipka 2017[13]	VAS score	1.204 (0.124 to 2.283)	0.190	1.275 (-0.115 to 2.665)	0.739	0.190
Mahon 1996[5]	Likert Scale (1-7)	0.069 (-0.042 to 0.179)	<0.001	0.145 (-0.153 to 0.443)	0.134	<0.001
Patel 1991[7]	4-item symptom questionnaire	0.000 (-0.000 to 0.000)*	<0.001	0.000 (-0.000 to 0.000)*	0.000	<0.001
Pereira 1995[8]	INR (diff)	0.027 (-0.155 to 0.209)	0.477	0.027 (-0.155 to 0.209)	0.000	0.477
Wallace 1994[9]	Conners 15-item rating scale scores	0.759 (0.341 to 1.178)	0.747	0.759 (0.341 to 1.178)	0.000	0.747
Woodfield 2005[10]	Number of cramps	-5.395 (-7.091 to -3.699)	<0.001	-18.823 (-28.527 to -9.120)	161.582	<0.001
Woodfield 2005[10]	Total days with cramps	-7.600 (-8.420 to -6.781)	<0.001	-6.181 (-9.798 to -2.563)	26.245	<0.001
Zucker 2006[11]	FIQ	-5.019 (-8.784 to -1.254)	0.999	-5.019 (-8.784 to -1.254)	0.000	0.999

* Includes one additional trial of Prednisone therapy

Appendix Table 6. Studies reporting person-level outcomes with both fixed-effect and random-effect hierarchical linear model

Author Year	Outcome	Range of the Scales (severity)	Fixed Treatment Effect	Random Treatment Effect	P-value Person Treatment Interaction
Camfield 1996[14]	Nights without awakening	NR	0.865 (0.215 to 1.516)	0.84 (0.20 to 1.48)	0.456
Hinderer 1990[15]	Anxiety	Beck Inventory-A anxiety scale 0-3 (0 = never, 3 = almost all the time)	0.000 (0.000 to 0.000)	-1.06 (-1.88 to -0.23)	<0.001
Joy 2014[26]	Myalgia score	Visual Analogue Score for myalgia (0=none to 100=worst)	3.3812 (-2.668 to 9.430)	3.3522 (-2.617 to 9.322)	0.566
Langer 1993[16]	Vomiting	NR	-1.204 (-2.494 to 0.086)	-1.20 (-2.49 to 0.09)	0.136
Lashner 1990[17]	Symptom score: abdominal pain	Symptom scores 0-100 (0=best, 100=worst)	-3.615 (-16.982 to 9.751)	-3.62 (-15.84 to 8.61)	0.007
	Symptom score: bowel movements/day		-0.538 (-1.215 to 0.138)	-0.56 (-1.22 to 0.09)	0.001
	Symptom score: consistency of bowel movements		7.000 (-7.551 to 21.551)	7.00 (-6.29 to 20.29)	0.013
	Symptom score: hematochezia		2.308 (-17.210 to 21.826)	2.35 (-17.21 to 21.90)	0.003
	Symptom score: general sense of well-being		-6.538 (-25.352 to 12.275)	-6.54 (-23.62 to 10.56)	0.008
Maier 1994[18]	SCL-90 subscales: Depressed mood	NR	-3.536 (-6.718 to -0.354)	-3.63 (-7.40 to 0.15)	<0.001
	SCL-90 subscales: Anxiety		-3.753 (-6.582 to -0.924)	-3.81 (-7.22 to -0.40)	<0.001
	SCL-90 subscales: Somatization		-1.419 (-4.316 to 1.478)	-1.50 (-4.20 to 1.21)	0.869
Mandelcorn 2004[19]	Self-Assessment score	0–5 (0=worst, 5=best)	-2.052 (-8.865 to 4.761)	-2.05 (-8.43 to 4.33)	0.05
	Lower extremity ataxia	Fugl-Meyer: 3-point (0 cannot be performed to 2 can	12.494 (-3.155 to 28.142)	12.49 (-0.85 to 25.84)	0.025

Author Year	Outcome	Range of the Scales (severity)	Fixed Treatment Effect	Random Treatment Effect	P-value Person Treatment Interaction
		be fully performed)			
	Truncal ataxia	AMTI forceplate®: NR *Berg Balance Scale® 0–56, with a higher score indicating a better performance*	1.196 (-2.866 to 5.257)	1.20 (-2.06 to 4.45)	0.690
	Upper extremity ataxia	Purdue Pegboard Test®: pegs inserted into the board with each hand in 30 sec *Minnesota Placing Test®: reach out, grasp, and place blocks in a specific order*	-0.498 (-3.546 to 2.550)	-0.50 (-3.10 to 2.10)	0.382
McQuay 1994[20]	VAS Pain Intensity	0-100 (0 = no pain, 100 = worst possible pain)	-1.094 (-5.572 to 3.383)	-1.06 (-5.16 to 3.04)	0.004
	VAS Relief Intensity	0-100 (0 = no relief, 100 =complete pain relief)	-3.913 (-11.729 to 3.903)	-3.86 (-11.11 to 3.40)	0.038
Miyazaki 1995[21]	Incidence of angina	Either ST-segment elevation or depression at rest	0.496 (-0.206 to 1.199)	0.47 (-0.32 to 1.26)	0.125
Nathan 2006[22]	Emetic episodes per day	complete response (0 episodes/day), major response (1–2 episodes/day), or failure (>2 episodes/day)	-0.095 (-0.514 to 0.325)	-0.56 (-1.74 to 0.62)	0.001
Parodi 1979[23]	Ischemic attacks	ST elevation or depression (details NR)	-1.544 (-1.838 to -1.251)	-1.63 (-2.10 to -1.17)	0.007
Parodi 1986[24]	Asymptomatic ST elevation (After verapamil)	NR	-1.637 (-1.994 to -1.279)	-1.97 (-2.92 to -1.01)	0.110
	Asymptomatic ST depression (After verapamil)		-1.083 (-1.903 to -0.262)	-0.82 (-2.54 to 0.90)	0.401

Author Year	Outcome	Range of the Scales (severity)	Fixed Treatment Effect	Random Treatment Effect	P-value Person Treatment Interaction
	Symptomatic ST elevation (After verapamil)		-1.580 (-1.906 to -1.254)	-1.87 (-2.72 to -1.02)	<0.001
	Symptomatic ST Depression (After verapamil)		-0.990 (-1.411 to -0.569)	-0.98 (-1.84 to -0.13)	0.002
	Asymptomatic ST elevation (After propranolol)		0.100 (-0.086 to 0.286)	-1.966 (-2.917 to -1.014)	0.006
	Asymptomatic ST depression (After propranolol)		0.339 (-0.168 to 0.845)	-0.821 (-2.539 to 0.897)	0.964
	Symptomatic ST elevation (After propranolol)		-0.002 (-0.177 to 0.173)	-1.868 (-2.718 to -1.017)	0.063
	Symptomatic ST Depression (After propranolol)		-0.374 (-0.709 to -0.039)	-0.981 (-1.835 to -0.126)	0.023
Pereira 1995[8]	INR	Target INR range of 2.0–3.0		-0.12 (-0.30 to 0.07)	0.433
Tison 2012[25]	Troublesome dyskinesia	7 points scale (1=extremely uncomfortable, 7=not at all uncomfortable)		0.20 (-0.40 to 0.80)	0.593

Statistical codes for analysis results of studies reporting person-level treatment effects

Estimation of standard errors in the following studies

- Emmanuel 2012: gen SE_Intervention (or control) =SD of intervention (or control) score/square root of Intervention days (or control days)
- Haas 2004: SE was available in Table 4 of the original paper
- Jaeschke 1991, Patel 1991, March 1994, Woodfield 2005, Wallace 1994 - SE was derived using the p-value of one-sided paired t-test of the difference in score using the following code:
  generate t_stat = invt(2,p_value)
  generate se = abs(mean_outcome/t_stat)
- Johannessen 1992, Pereira 1995, Zucker 2006, Joy 2014, Lipka 2017 – SE was derived from the 95% confidence interval using the following code: generate se = (UCI - LCI) /(2*invnorm(0.975))
- Mahon 1996: SE was derived from 95% confidence interval based on Student's t distribution using the following code: generate se = (UCI - LCI) /(2*invt(DF, 0.975))

```
metan difference se_difference  if Outcome == "outcome", random        **/fixedi is used for fixed effect model

    local p = r(p_het)

    local sum_es =  r(ES)

    local sum_es_se = r(seES)

    local tau2=  r(tau2)

    local I_sq =  r(i_sq)

post `memory' ("`study'") ("`outcome'")  (`sum_es') (`sum_es_se')  (`tau2') (`I_sq') (`p')
```

Statistical codes for analysis results of studies reporting person-level outcome effects

```
egen id = group(Patient)
generate tx = 0 if Exposure == "Placebo"
        replace tx = 1 if Exposure == "Intervention"
egen period_seq = seq(), from(1) to(18) */varies based on the number of periods*/
local outcome = "Specific_outcome"
        /* fixed baselines and random treatment effects */
        xtmixed Result  tx i.id || id: tx  if Outcome == "`outcome'"  , nocons
                estimates store D
                matrix estimates = e(b)
                local point_estimate_ran_bas_ran_tx = estimates[1,1]
                local sd_estimate_rand_base_random_tx = (exp(estimates[1,10]))

                matrix variances = e(V)
                local point_se_rand_base_random_tx = sqrt(variances[1,1])
                local point_low_ran_bas_ran_tx = `point_estimate_ran_bas_ran_tx' - invnormal(0.975) * `point_se_rand_base_random_tx'
                local point_up_ran_bas_ran_tx = `point_estimate_ran_bas_ran_tx' + invnormal(0.975) * `point_se_rand_base_random_tx'

                local sd_se_rand_base_random_tx = sqrt(variances[10,10])
```

```
		local sd_lower_rand_base_random_tx = (exp(ln((`sd_estimate_rand_base_random_tx')) - invnormal(0.975) *
`sd_se_rand_base_random_tx'))

		local sd_upper_rand_base_random_tx = (exp(ln((`sd_estimate_rand_base_random_tx')) + invnormal(0.975) *
`sd_se_rand_base_random_tx'))


	/* fixed baselines and common treatment effect -- linear regression */

	xtmixed Result  tx i.id  || id: if Outcome == "`outcome'" , nocons

		estimates store E


	/* fixed baselines and person interactions */

	regress Result i.tx##i.id if Outcome == "`outcome'"

		estimates store F


	/* fixed baselines and common effects */

	regress Result tx i.id if Outcome == "`outcome'"

		estimates store G


		matrix estimates = e(b)

		local point_estimate_fix_bas_com_tx = estimates[1,1]
```

```
            matrix variances = e(V)
            local point_se_fix_bas_common_tx = sqrt(variances[1,1])
            local t_stat = `point_estimate_fix_bas_com_tx' /  `point_se_fix_bas_common_tx'
            local point_low_fix_bas_com_tx = `point_estimate_fix_bas_com_tx' - invt(e(df_r), 0.975) * `point_se_fix_bas_common_tx'
            local point_up_fix_bas_com_tx = `point_estimate_fix_bas_com_tx' + invt(e(df_r), 0.975) * `point_se_fix_bas_common_tx'

        lrtest D E
            local p_random_RANDOM_FIXED_tx =  r(p)

        lrtest F G
            local p_person_by_treat =  r(p)

        post `memory'   ("Study") ("`outcome'")
```

Please note: Depending on the outcome, xtmixed or meqrlogit or meqrpoisson was used.

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	1-2
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	2
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	n/a
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5-6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5-6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	a1-a3
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	6-7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	6-8
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	n/a
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	8-9
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	8-9

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	n/a
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	8-9
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	9, 20, 21, 29
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	1-12, 22-26
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	n/a
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	31, a11-a50
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	10-12, 26
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	n/a
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	12, a53-a57
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	12-14
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	15-16
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	16
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	17

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

PRISMA 2009 Checklist