Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Power Simulation Guide for Choosing Longitudinal Endpoints and Analysis Methods for Continuous Data

Paul J. Novotny

novotny@mayo.edu

Mayo Clinic

Charles L. Loprinzi

Mayo Clinic

Jennifer G. Le-Rademacher

Mayo Clinic

Article

Keywords: longitudinal, endpoint, simulation, copula, power

Posted Date: March 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8845743/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background: The choice of analysis method and primary endpoint can have a major impact on the results of a clinical trial. The power of these analysis methods is affected by the response patterns, response distributions, autocorrelations, missing values, and the number of times the endpoint is evaluated. Guidelines are needed to help investigators select statistical methods that give investigators the best chance of finding an intervention effect if one truly exists.

Methods: Monte Carlo methods were used to determine how different endpoints are affected by changes in study assumptions. Copula simulation methods were applied to prior Alliance hot flash and CIPN studies to determine which endpoints would provide the highest power for future studies.

Results: Longitudinal analyses had higher power if there was a plateau response, there were fewer missing values, responses were collected more often, and there was lower correlation between responses from the same subject. Having skewed data did not greatly affect power. Prior hot flash and CIPN studies showed skewed data, plateau response rates, and very high autocorrelations.

Conclusion: Hot flash and CIPN studies are more likely to find true effects if the analysis method uses data at all of the time points and adjusts for baseline. Collecting data at more time points and reducing the missing data rates also increase study power.

Plain English Summary

The power of analyses of repeated measures over time is affected by the choice of the analysis method and by characteristics of the data. Power is higher if responses are collected at more time points, the analysis method uses data at all time points, there are fewer missing values, responses from the same subject have low correlations, and if responses have a plateau effect (having a maximum change at early time points followed by a flat response after that).

BACKGROUND

Patient-reported outcomes are important benchmarks to understand the toxicities and symptoms that cancer patient experienced as a result of their cancer or their cancer therapy. Common symptoms include chemotherapy-induced neuropathy, chemotherapy-induced nausea and vomiting, hot flashes, mucositis, skin rash, etc. Although many of these symptoms can be reported and their severity can be graded by clinicians using tools such as the National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events [1], these symptoms are often burdensome to patients and have significant adverse effects on patient quality of life. For these reasons, it is prudent to use patient-centered outcomes as the primary endpoint when evaluating the effectiveness of interventions that aim to alleviate symptoms associated with cancer therapy. Patient-reported symptom severity and/or level of interference are referred to hereafter as symptom score for brevity.

In symptom intervention trials with Patient Reported Outcomes (PROs), patients are asked to provide their assessment of the symptom under investigation, via questionnaires, at pre-specified time points. In most trials, the time points include a baseline assessment that occurs just prior to starting the intervention and subsequent time points that occur during treatment and follow-up. Since the level of sensitivity to a symptom may vary between patients, the baseline assessment acts as an anchor so that the effect of an intervention can be judged as an

increase or a decrease in symptom severity and/or interference after starting the intervention relative to their baseline level.

Choosing the appropriate primary endpoint and the corresponding analysis method is one of the most essential, but often challenging, aspects of clinical trial design. For trials with PROs, the choice of the primary endpoint and analysis method is even more challenging due to the large number of endpoints that can be derived from the repeated assessments and corresponding analysis methods. Choice of endpoints can range from using the scores at each time point to using just the final symptom score. Choices of analysis can range from running t-tests at a specific time point to repeated measures mixed models that account for all of the time points and the correlation between measurements [2], Note that although continuous data can be transformed into categorical data or time-to-event data, this work focuses on endpoints and analyses where the PROs are preserved as continuous variables.

Although there is vast literature on design and analysis of longitudinal data, there are few guidelines for choosing the primary endpoint and the analysis method in clinical trials that use continuous longitudinal PRO data. The choice of the preferred endpoint and appropriate analysis method depend on the expected response patterns, the expected response distribution, the correlation between consecutive measurements, and the amount and structure of any missing data [3]. There are few guidelines for when to choose one endpoint over another. This study used Monte Carlo simulation to evaluate the statistical power of different endpoints and analysis methods under a variety of assumptions about the true nature of the responses and the pattern of missing values. The knowledge gained from the simulation study will be applied to data from completed Alliance for Clinical Trials in Oncology (Alliance) symptom intervention trials to provide guidance for selecting the most appropriate endpoint and statistical methods in future trials in these settings. The results from this study will further be formulated as guidelines for selection of endpoints and analysis methods for trials beyond the symptom intervention setting with PROs.

Patient-reported outcomes are used in a large number of clinical trials, especially in symptom intervention trials, other cancer control trials and, and trials with Quality of Life (QOL) endpoints. The primary endpoint formulations and corresponding analysis methods have often been selected in an ad-hoc manner based on the study team's experience and judgement. This current report provides evidence-based recommendations for selection of optimal endpoints and analyses for future clinical trials with PROs. This will, in turn, improve the statistical strengths of study results and increase confidence in patient-reported outcomes. The recommended optimal endpoints and analysis methods are crucially important for determining sample size when designing trials that use continuous data collected at multiple time points.

METHODS

This is a Monte Carlo simulation study to determine the power of longitudinal study endpoints and analysis methods for continuous responses. The simulations look at how power is affected by assumptions concerning response patterns, response distributions, autocorrelations, missing data patterns, and the number of responses per subject. To allow for small confidence intervals around the power estimates [4] and to control computer execution time, ten thousand simulated datasets were generated for each combination of assumptions. Each analysis method was applied to the simulated data to obtain the proportion of times the tests rejected the null hypothesis. These proportions were combined to estimate the power, with 95% confidence intervals for each method [5,6]. All of the tests were two-sided comparing two arms. The simulations and analyses for this paper were generated using SAS software. Copyright © 2016 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

Statistical Methods

Data simulation

Three simulation methods were used in this study[6,7]. In the first simulation method, data at each time point was simulated as a linear function of the data at the previous time point. Without loss of generality, the simulations were based on 5 evenly spaced time points with the responses denoted by Y_i for i=0, 1, 2, 3, and 4 corresponding to baseline and four time points post baseline. At baseline, values Y_0 were simulated from a normal distribution with a mean of 50 and a standard deviation of 10. These are common assumptions for PROs such as the SF-36 [8] and PROMIS [9]. Subsequent values were simulated as $Y_{i+1}=Y_0+ C_i*\delta +e$, where e is a random deviate from a normal distribution with a mean of 0 and a standard deviation of 10, C_i is a coefficient based on the response pattern (see Table 1 in the Response Pattern section below), and δ is the maximum level of change from baseline (2 for a small change, 5 for a moderate change, 8 for a large change, and 10 for a one standard deviation change).

The second simulation method used SAS proc simnormal to simulated correlated normally distributed data. The third method involved using copula methods to simulate correlated uniformly distributed data and then transforming these uniform distributions to other distributions. To simulate skewed data, these uniform values were transformed to beta distributions over a restricted range.

Copula simulation [10,11] was also applied directly to data collected on Alliance hot-flash and chemotherapy-induced peripheral neuropathy (CIPN) clinical trials. Copulas can simulate data with any correlation structure and any marginal distributions at each time point. It starts by creating uniform distributions that have your desired correlation structure and transforms the uniform distributions to the desired marginal distributions. This allows you to simulate data with the correlation structure and mariginal distributions like those seen in prior studies. These copula methods were used to validate the results from the other simulation methods.

Response Patterns

In the first two simulation methods, assumptions were made about the expected response trends over time. Five different alternatives were explored, as illustrated in Figure 1. A linear response simulates a straight line change over time, a delayed response simulates having no immediate response but a linear response later in time, a temporary response simulates an early response followed by a regression back to the starting level, a plateau change simulates an early change in responses followed by no further change, and the final response pattern was having no change over time. Each of these response categories were simulated for small (2 point), medium (5 point), large (8 point) and one standard deviation (10 point) effect size changes [12].

Table 1: C_i constants

Response Pattern	C_0	C_1	C_2	C_3
No Change	0	0	0	0
Linear Increase	0.25	0.5	0.75	1
Delayed Increase	0	0	0.5	1
Temporary Increase	0.5	1	0.5	0
Plateau Increase	0.5	1	1	1

Correlations between responses from the same subject

Responses from the same subject had one of these two assumptions: 1) responses were not correlated or 2) responses followed a first-order autoregressive structure, AR(1). An AR(1) model assumes responses from the same subject at different time points are correlated, but the correlation is smaller if the time points are farther apart. For example, an AR(1) with an autocorrelation of 0.5 assumes responses one time point apart have a correlation of 0.5, responses two time points apart had a correlation of $(0.5)^2$=0.25, responses three time points apart had a correlation of $(0.5)^3$=0.125, and responses four time points apart had a correlation of $(0.5)^4$=0.0625.

Missing values

Analyses were first done assuming there were no missing values. Additional analyses assumed missing values were either missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR) [13]. The proportion of missing values ranged from 0% to 50%.

MCAR means the missing values are missing completely at random and are not related to any study values. MCAR missing values were simulated by giving each response at each time point a fixed likelihood of being missing. For example, if the missing rate is set to 10%, each simulated value was given a 10% chance of being missing.

MAR means the missing values are related to observed study variables. MAR missing values were simulated by setting the likelihood of a value being missing conditional on the previous reported value.

MNAR means the likelihood of a value being missing is related to the actual value. For example, low PRO values might be more likely to be missing. MNAR missing values were simulated by setting the likelihood of a value being missing conditional on the level of the value. Work by Mazza [14] provides more details about these simulation methods for missing values.

Endpoint Analysis Methods

There are numerous ways of analyzing longitudinal data[15]: GEE modeling [16,17], linear growth mixed models, responder analyses (whether or not the subject obtained a specified level of response), the time until a response was observed, AUC (area under the curve), AUC adjusted for baseline value, the average of scores over all time points, the maximum (or minimum) value obtained, individual patient slopes over time, changes from baseline to last value, the percentage change from baseline to last value, or just the response at the last time point. The analysis methods are usually based on t-tests, but nonparametric tests (like Wilcoxon rank sum tests) are often used for non-normally distributed data. There are also many different methods of adjusting for baseline values. While all of these methods were explored in the course of this study, this manuscript reports on the methods that are most often used in clinical trials: comparing change from baseline to the end of the study using t-tests, growth curve mixed models, t-tests of the area under the curve, t-tests of the average over all time points, t-tests of the slope of the scores over time for each patient, responder analyses, and GEE modeling.

Monte Carlo Assumptions

Ten thousand simulations were created for each combination of assumptions about the response patterns in the control group, response patterns in the intervention group, response distributions, correlations between repeated measures on the same subject, the number of time points, and assumptions about missing values. For each simulation, all of the different analysis methods were applied to the data to determine whether there was a significant difference between the intervention and control arms at an alpha level of 0.05 with two-sided testing. The proportion of the 10,000 simulations that resulted in a significant difference was used as the power of the analysis method. The 2.5^{th} and 97.5^{th} percentiles were used as the confidence intervals for the power [5].

Copula simulation methods [10] were applied to completed Alliance NCI Community Oncology Research Program (NCORP) symptom intervention studies to validate the recommendations for the optimal statistical methods for these types of studies. Data from completed CIPN and hot flash studies were used to better understand their response distributions, response patterns, correlations between responses from the same subject, and the percent of missing values. Using the observed characteristics based on this real data, simulations were run to estimate the power from each analysis method. Separate determinations were made for CIPN studies and for hot flash studies to help select the analysis method with the highest power for future studies.

Patients

Data was collected from 7 CIPN and 14 hot flash studies (table 7) run through the Alliance for Clinical Trials. These studies were selected, because they were phase III studies with both control and intervention arms. These studies were reviewed by the Institutional Review Board (IRB) of each of the participating sites, and all patients provided informed consent to participate in the study. These hot-flash and CIPN studies were performed in line with the principles of the Declaration of Helsinki and approved by the Alliance ethics committee.

Measures

For CIPN studies, data was collected using the European Organisation For Research and Treatment of Cancer (EORTC) CIPN-20 [18]. For hot flash studies, data was collected using self-report diaries [19].

RESULTS

All of the results are based on the following assumptions unless specified otherwise: 10,000 simulations, 50 subjects in each arm, data collected at 5 time points, responses with a normal distribution with a mean of 50 and standard deviation of 10 at baseline, no correlation between repeated measure on the same person, the control arm had no change, the intervention arm had a linear response pattern with a maximum increase of 5 points over time, 5% of data were MAR, data was set to MCAR if the previous value was in the lower 5% of the assumed distribution, and data was set to MNAR if the current value was in the lower 5% of the assumed distribution. The sample size of 50 was selected so there would be reasonable power differences between endpoints. Larger sample sizes would result in all of the endpoints having power close to one, and smaller samples would result in all of the endpoints having power close to zero.

Effects of Response Pattern

Table 2 shows the power (95% confidence interval) for each study endpoint when response patterns were changed in the intervention arm. Plateau changes were the easiest to detect with all endpoints having the highest power for this response pattern. Temporary and delayed responses were the hardest to detect. Patient slopes and change from baseline had almost no power to detect temporary changes.

Effects of the Number of Time Points Measured

Table 3 shows power (95% confidence interval) for each study endpoint if the data were collected at 2, 3, 4, 5, 10, or 20 different time points. All study endpoints (except for the change from baseline) had higher power if the data was collected more often.

Effects of Missing Values

Table 4 shows power (95% confidence interval) for each study endpoints under different assumptions about the rate of missing values. For all endpoints, power decreased as the rate of missing values increased. The AUC showed the biggest drop in power as the rate of missing values increased.

Effects of Correlation Between Repeated Measures on the Same Subject

Table 5 shows the power (95% confidence interval) for each study endpoint under different assumptions about the correlation between repeated measures. These simulations assume the correlation is autoregressive. The power for most of the endpoints went down as the correlations increased. However, the power for the change from baseline and patient slopes increased as the correlations increased.

Effects of Non-Normal Data

To determine the effects of having skewed data, data was simulated using two different methods. Transformed beta distributions were used to simulate skewed data. Skewed left data was simulated by setting the maximum possible value close to the mean and had a skewness around -1.0. Skewed right data was simulated by setting the minimum possible value close to the mean and had a skewness around 1.0. The beta distributions were transformed from their original range from 0 to 1 so that the data had a mean of 50 and a standard deviation of 10 at baseline. Skewed data are often transformed used square root or log transformations to make the data look more normally distributed, so the power based on those transformations are also included. Table 6 shows power (95% confidence interval) for each study endpoint for these transformed results. The Wilcoxon tests performed similar to the t-tests. Power was not greatly affected by the skewness of the distribution of responses and transforming the skewed data also did not greatly affect power.

Alliance CIPN and Hot Flash Copula Results

Study Demographics

Table 7 shows the characteristics for each of the hot flash and CIPN studies. Correlations stayed very high even for measurements that were separated by several time points.

The power of a study is affected by a combination of these study characteristics. Copulas provide a framework for understanding the combined effects of the study characteristics on power. Copulas were used to simulate two sets of responses for each replication: a set similar to the intervention arm and another set similar to the control arm. Tables 8 and 9 show the power for each endpoint applied to simulations based on these real CIPN and hot flash data. The endpoint with the highest power varied by study. There was no single endpoint that was clearly better across all studies. Of interest, there are studies, like the citalopram study, where it was essential to adjust for baseline values. This is because the treatment effect occurred in the first time point after baseline, and there were no additional changes after the first time point. Without adjusting for baseline values, the responses on the treatment and control arms looked very similar. Responder analyses had very low power for these studies.

DISCUSSION

The choice of analysis method and primary endpoint has a major impact on whether an intervention is determined to be effective. The power of these analyses methods is greatly affected by the response patterns in the control and intervention arms, the correlation between repeated measurements on the same subject, the type and extent of missing values, the distribution of the responses, and the number of times the endpoint is evaluated over time. Power was higher for almost all study endpoints if there was a plateau response effect, meaning the full effect happens

early and is maintained until the end of the study. Endpoints that use only a small number of time points, like the change from baseline, had low power if the response patterns showed only a temporary change or a delayed change. There was higher power if there were fewer missing values, responses are collected more often, and there was a lower correlation between responses from the same subject.

Alliance hot flash studies usually had plateau response patterns in both the control and intervention arms, had about 11% of the responses missing, had skewed data, and had very high autocorrelations (around 0.9). Response patterns for CIPN varied, with many of the studies showing no change, and other studies showing linear or plateau responses.

The main limitation of this study was that these analyses only looked at changing one aspect of the simulations at a time, such as changing only the autocorrelation assumptions. It is likely that power could change differently if combinations of the responses changed and if the sample sizes were different. For example, if autocorrelations changed along with the missing data rates, there could be an interaction between these two assumptions. Copula simulation was able to make combinations of assumptions that most closely correspond to characteristics observed on prior studies.

CONCLUSION

The choice of study endpoints and analysis methods is affected by many aspects of the study design. If there is no treatment effect, it does not matter what endpoint or analysis method is used. For large treatment effects, any endpoint will detect the effect. If there is only a temporary difference between groups, endpoints that use all time points, such as AUC and linear growth models, are more likely to find the effect.

Prior hot flash and CIPN studies showed varied response patterns, with plateau responses being the most prominent. They also had very high autocorrelations. Missing data rates were at an acceptable level which resulted in minimal impact on the study power. The characteristics of these studies suggest that analysis methods that use all observed data, such as AUC analysis or using the sum of the responses across all time points, would result in the highest power for hot flash and CIPN studies. It was also essential to include an adjustment for baseline values in these studies. This is because treatment effects occurred quickly and plateaued. Without adjusting for baseline values, there could appear to be no difference between arms.

It is not possible to look at how power changes for all possible combinations of assumptions about a study. Future work includes development of an interactive application where a researcher can enter assumptions about baseline distributions, underlying response patterns, response distributions, correlation between repeated measures on the same subject, missing data rates, missing data types, the number of times the endpoint is assessed, and the effect size. The user could enter a sample size and receive estimates of the power for each endpoint and analysis method under these assumptions. They could then look at a range of sample sizes to obtain a desired power level. This will guide researchers in choosing a primary endpoint that would have the highest likelihood of finding a significant difference if there is a true intervention effect [38].

Abbreviations

NCORP = NCI Community Oncology Research Program

PRO = Patient Reported Outcomes

PRO-CTCAE = Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events

SD = Standard Deviation

Declarations

Ethical Approval and Consent to participate: The study was reviewed by the IRB of each of the participating sites, and all patients provided consent to participate in their original studies.

Consent for publication: Not applicable

Availability of data and materials: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request. The simulation code is available on request from the corresponding author.

Competing interests: The authors have no conflicts of interest.

Funding: This study was supported by the Daniel J. Sargent, PhD Memorial Fellowship in Innovative Clinical Trial Design and Methods.

Author Contributions: All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Paul J. Novotny and Jennifer G. Le-Rademacher. The first draft of the manuscript was written by Paul J. Novotny and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Acknowledgements: Not applicable

References

1. U.S. Department of Health and Human Services (2017), Common Terminology Criteria for Adverse Events (CTCAE) Version 5.0, November 27, https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/CTCAE_v5_Quick_Reference_5x7.pdf
2. Le-Rademacher, J.G., Storrick, E.M. and Jatoi, A. (2019). Remarks on the design and analyses of longitudinal studies for cancer patients with anorexia and weight loss. *Journal of cachexia, sarcopenia and muscle*, 10(6), pp.1175-1182.
3. Touloumi, G., Babiker, A.G., Pocock, S.J. and Darbyshire, J.H. (2001). Impact of missing data due to drop-outs on estimators for rates of change in longitudinal studies: a simulation study. *Statistics in medicine*, 20(24), pp.3715-3728.
4. Morris, T.P., White, I.R. and Crowther, M.J., 2019. Using simulation studies to evaluate statistical methods. Statistics in medicine, 38(11), pp.2074-2102.
5. Efron B (1982), The jackknife, the bootstrap and other resampling plans, Regional Conference Series in Applied Mathematics, 38. SIAM, Philadelphia.
6. Wicklin, Rick (2015). "Ten Tips for Simulating Data with SAS®." *SAS Global Forum Proceedings*. https://support. sas. com/resources/papers/proceedings15/SAS1387-2015. pdf (Paper SAS1387-2015).
7. Wicklin, Rick (2013). *Simulating data with SAS*. SAS Institute.
8. Ware Jr, J. E. (1999). SF-36 health survey.
9. Cella, D., Riley, W., Stone, A. A., Rothrock, N., Reeve, B. B., Yount, S., Amtmann, D., D., B., Choi, S., Cook, K. F., et al. (2010). The Patient Reported Outcomes Measurement Information

10. System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *Journal of Clinical Epidemiology* 63, 1179-1194.
11. Nelsen, R.B. (2007). *An introduction to copulas*. Springer Science & Business Media.
12. Trivedi PK, Zimmer DM (2007). *Copula modeling: an introduction for practitioners*. Now Publishers Inc.
13. Sloan, J.A., Cella, D. and Hays, R.D. (2005). Clinical significance of patient-reported questionnaire data: another step toward consensus. *Journal of clinical epidemiology*, 58(12), pp.1217-1219.
14. Fairclough DL (2004) *Design and Analysis of Quality of Life Studies in Clinical Trials*, CRC Press.
15. Mazza, G.L., Kunze, K.L., DeWees, T.A. and Dueck, A.C. (2019), October. Monte Carlo simulations evaluating the impact of missing data on between-arm comparisons on summary measures in randomized trials. *Quality of Life Research* (Vol. 28, pp. S74-S75). VAN Godewijckstraat 30, 3311 Gz Dordrecht, Netherlands: Springer.
16. Newsom, J.T. (2012). Basic longitudinal analysis approaches for continuous and categorical variables. *Longitudinal data analysis: A practical guide for researchers in aging, health, and social sciences*, pp.143-179.
17. Diggle, P. J., Liang, K. Y. and Zeger, S. L. (1993) *Analysis of Longitudinal Data*, Oxford University Press,Oxford.
18. Liang, K.Y. and Zeger, S.L. (2000). Longitudinal data analysis of continuous and discrete responses for pre-post designs. Sankhyā: *The Indian Journal of Statistics*, Series B, pp.134-148.
19. Postma, T.J., Aaronson, N.K., Heimans, J.J., Muller, M.J., Hildebrand, J.G., Delattre, J.Y., Hoang-Xuan, K., Lantéri-Minet, M., Grant, R., Huddart, R. and Moynihan, C. (2005). The development of an EORTC quality of life questionnaire to assess chemotherapy-induced peripheral neuropathy: the QLQ-CIPN20. *European journal of cancer*, 41(8), pp.1135-1139.
20. Sloan, J.A., Loprinzi, C.L., Novotny, P.J., Barton, D.L., Lavasseur, B.I. and Windschitl, H. (2001). Methodologic lessons learned from hot flash studies. *Journal of Clinical Oncology*, 19(23), pp.4280-4290.
21. Barton, D.L., Loprinzi, C.L., Quella, S.K., Sloan, J.A., Veeder, M.H., Egner, J.R., Fidler, P., Stella, P.J., Swan, D.K., Vaught, N.L. and Novotny, P. (1998). Prospective evaluation of vitamin E for hot flashes in breast cancer survivors. *Journal of Clinical Oncology*, 16(2), pp.495-500.
22. Pruthi, S., Qin, R., Terstreip, S.A., Liu, H., Loprinzi, C.L., Shah, T.R., Tucker, K.F., Dakhil, S.R., Bury, M.J., Carolla, R.L. and Steen, P.D. (2012). A phase III, randomized, placebo-controlled, double-blind trial of flaxseed for the treatment of hot flashes: NCCTG N08C7. *Menopause* (New York, NY), 19(1), p.48.
23. Loprinzi, C.L., Kugler, J.W., Sloan, J.A., Mailliard, J.A., LaVasseur, B.I., Barton, D.L., Novotny, P.J., Dakhil, S.R., Rodger, K., Rummans, T.A. and Christensen, B.J. (2000). Venlafaxine in management of hot flashes in survivors of breast cancer: a randomised controlled trial. *The Lancet*, 356(9247), pp.2059-2063.
24. Barton, Debra L., Beth I. LaVasseur, Jeff A. Sloan, Allen N. Stawis, Kathleen A. Flynn, Missy Dyar, David B. Johnson, Pamela J. Atherton, Brent Diekmann, and Charles L. Loprinzi. (2010) "Phase III, placebo-controlled trial of three doses of citalopram for the treatment of hot flashes: NCCTG trial N05C9." *Journal of Clinical Oncology* 28, no. 20: 3278.
25. Goldberg, Richard M., Charles L. Loprinzi, Judith R. O'Fallon, Michael H. Veeder, Angela W. Miser, James A. Mailliard, John C. Michalak, Ann Marie Dose, Ken M. Rowland Jr, and Nora L. Burnham (1996). "Transdermal clonidine for ameliorating tamoxifen-induced hot flashes." *Journal of Clinical Oncology* 14, no. 8: 2411-2411.
26. Pockaj, B.A., Gallagher, J.G., Loprinzi, C.L., Stella, P.J., Barton, D.L., Sloan, J.A., Lavasseur, B.I., Rao, R.M., Fitch, T.R., Rowland, K.M. and Novotny, P.J. (2006). Phase III double-blind, randomized, placebo-controlled crossover trial of black cohash in the management of hot flashes: NCCTG Trial N01CC1. *J Clin Oncol*, 24(18), pp.2836-2841.
27. Loprinzi, C.L., Michalak, J.C., Quella, S.K., O'fallon, J.R., Hatfield, A.K., Nelimark, R.A., Dose, A.M., Fischer, T., Johnson, C., Klatt, N.E. and Bate, W.W. (1994). Megestrol acetate for the prevention of hot flashes. *New England*

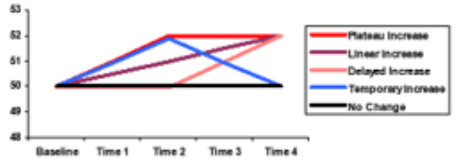
Journal of Medicine, 331(6), pp.347-352.

28. Park, H., Qin, R., Smith, T.J., Atherton, P.J., Barton, D.L., Sturtz, K., Dakhil, S.R., Anderson, D.M., Flynn, K., Puttabasavaiah, S. and Le-Lindqwister, N.A. (2015). North Central Cancer Treatment Group N10C2 (Alliance): a double-blind placebo-controlled study of magnesium supplements to reduce menopausal hot flashes. *Menopause*, 22(6), pp.627-632.
29. Leon-Ferre, R.A., Novotny, P.J., Wolfe, E.G., Faubion, S.S., Ruddy, K.J., Flora, D., Dakhil, C.S., Rowland, K.M., Graham, M.L., Le-Lindqwister, N. and Smith, T.J. (2020). Oxybutynin vs placebo for hot flashes in women with or without breast cancer: a randomized, double-blind clinical trial (ACCRU SC-1603). *JNCI Cancer Spectrum*, 4(1), p.pkz088.
30. Loprinzi, C.L., Qin, R., Baclueva, E.P., Flynn, K.A., Rowland Jr, K.M., Graham, D.L., Erwin, N.K., Dakhil, S.R., Jurgens, D.J. and Burger, K.N. (2010). Phase III, randomized, double-blind, placebo-controlled evaluation of pregabalin for alleviating hot flashes, N07C1. *Journal of Clinical Oncology*, 28(4), p.641.
31. Loprinzi, C.L., Sloan, J.A., Perez, E.A., Quella, S.K., Stella, P.J., Mailliard, J.A., Halyard, M.Y., Pruthi, S., Novotny, P.J. and Rummans, T.A. (2002). Phase III evaluation of fluoxetine for treatment of hot flashes. *Journal of Clinical Oncology*, 20(6), pp.1578-1583.
32. Quella, S.K., Loprinzi, C.L., Barton, D.L., Knost, J.A., Sloan, J.A., LaVasseur, B.I., Swan, D., Krupp, K.R., Miller, K.D. and Novotny, P.J. (2000). Evaluation of soy phytoestrogens for the treatment of hot flashes in breast cancer survivors: A North Central Cancer Treatment Group Trial. *Journal of Clinical Oncology*, 18(5), pp.1068-1068.
33. Rao, R.D., Michalak, J.C., Sloan, J.A., Loprinzi, C.L., Soori, G.S., Nikcevich, D.A., Warner, D.O., Novotny, P., Kutteh, L.A., Wong, G.Y. and North Central Cancer Treatment Group (2007). Efficacy of gabapentin in the management of chemotherapy-induced peripheral neuropathy: a phase 3 randomized, double-blind, placebo-controlled, crossover trial (N00C3). *Cancer*, 110(9), pp.2110-2118.
34. Rao, R.D., Flynn, P.J., Sloan, J.A., Wong, G.Y., Novotny, P., Johnson, D.B., Gross, H.M., Renno, S.I., Nashawaty, M. and Loprinzi, C.L. (2008). Efficacy of lamotrigine in the management of chemotherapy-induced peripheral neuropathy: A phase 3 randomized, double-blind, placebo-controlled trial, N01C3. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 112(12), pp.2802-2808.
35. Grothey A, Nikcevich DA, Sloan JA, Kugler JW, Silberstein PT, Dentchev T, Wender DB, Novotny PJ, Chitaley U, Alberts SR, Loprinzi CL (2011) Intravenous calcium and magnesium for oxaliplatin-induced sensory neurotoxicity in adjuvant colon cancer: NCCTG N04C7. *J Clin Oncol* 29:421–427.
36. Kottschade, L.A., Sloan, J.A., Mazurczak, M.A., Johnson, D.B., Murphy, B., Rowland, K.M., Smith, D.A., Berg, A., Stella, P.J. and Loprinzi, C.L. (2009). The use of vitamin E for prevention of chemotherapy-induced peripheral neuropathy: A phase III double-blind, placebo controlled study—N05C31. *Journal of Clinical Oncology*, 27(15_suppl), pp.9532-9532.
37. Barton, D.L., Wos, E., Qin, R., Mattar, B., Green, N., Lanier, K., Bearden, J., Kugler, J., Rowland Jr, K. and Loprinzi, C. (2009). A randomized controlled trial evaluating a topical treatment for chemotherapy-induced neuropathy: NCCTG trial N06CA. *Journal of Clinical Oncology*, 27(15_suppl), pp.9531-9531.
38. Leal, A.D., Qin, R., Atherton, P.J., Haluska Jr, P., Behrens, R.J., Tiber, C., Watanaboonyakhet, P., Weiss, M., Adams, P.T., Dockter, T. and Loprinzi, C.L. (2014). NCCTG N08CA (Alliance): The use of glutathione for prevention of paclitaxel/carboplatin induced peripheral neuropathy: A phase III randomized, double-blind placebo-controlled study. *Cancer*, 120(12), p.1890.
39. Loprinzi, C.L., Qin, R., Dakhil, S.R., Fehrenbacher, L., Flynn, K.A., Atherton, P., Seisler, D., Qamar, R., Lewis, G.C. and Grothey, A. (2014). Phase III randomized, placebo-controlled, double-blind study of intravenous calcium and magnesium to prevent oxaliplatin-induced sensory neurotoxicity (N08CB/Alliance). *Journal of Clinical Oncology*, 32(10), p.997.

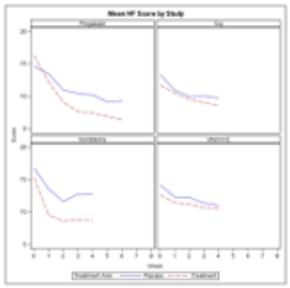
Tables

Tables 1 to 9 are available in the Supplementary Files section.

Figures



Plots of mean values over time for CIPN studies.



Plots of mean values over time for CIPN studies.

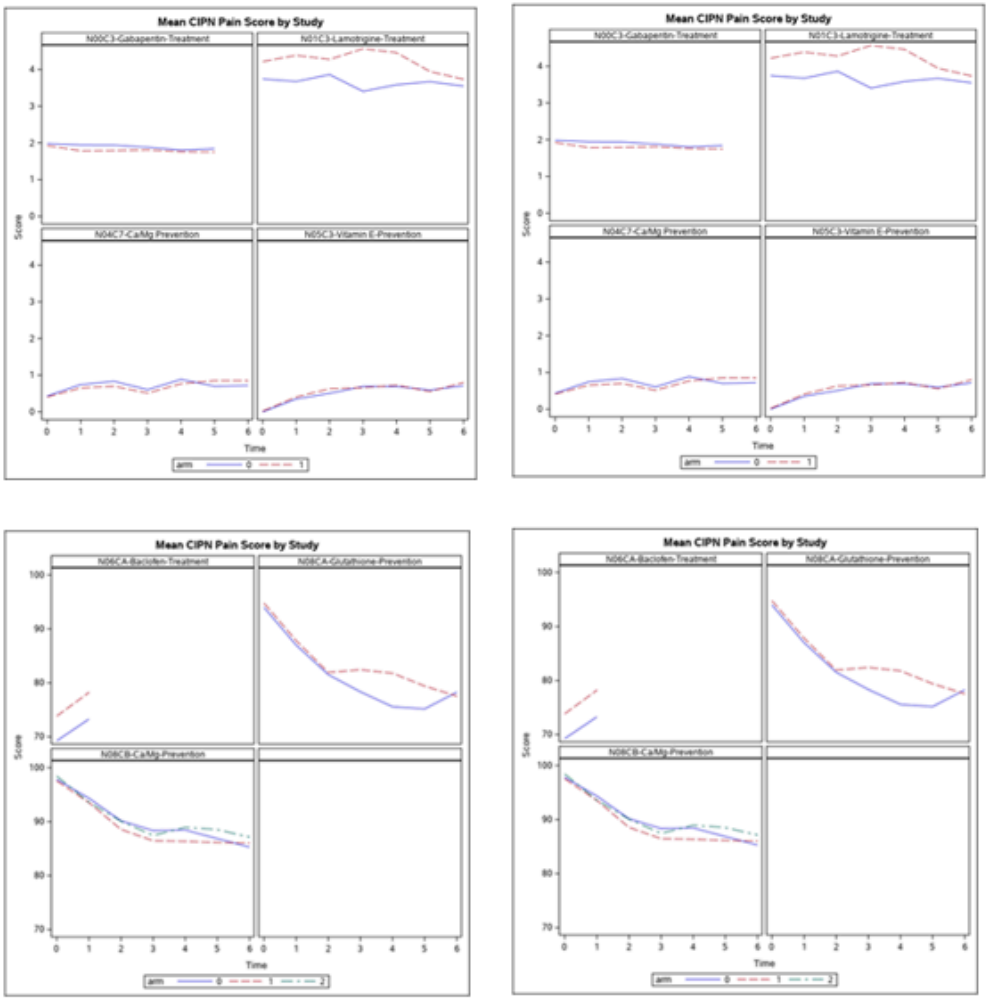


Figure 1

Response Patterns

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Tables.docx